Biological Responses to Therapeutic Treatments of Human Vascular Diseases

Thesis by

Zixuan Shao

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2019

(Defended December 21, 2018)

Zixuan Shao

ORCID: 0000-0002-4676-6023

Dedicated to my Family

Acknowledgements

There are many people I want to thank for helping me throughout my graduate study. Without their academic and personal support, I would not have been able to survive the last five and half years of rigorous scientific exploration.

The first person I would like to thank is my advisor, Julie Kornfield. First and foremost, Julie taught me to be quantitative and rigorous in my research – a switch from not only thinking like a biologist but also as an engineer. In particular, the high bar Julie sets for her students helped me learn how to ask better scientific questions and how to carry out more reproducible experiments. During my study, Julie also helped me make substantial improvements on my interpersonal skills. Learning from her example, I became a better communicator of science and a more supportive colleague. And without her help, we would had a much harder time navigating the numerous collaborations that became an integral part of my research. Thank you, Julie, for making me a better scientist and a better person, and for creating a supportive environment for doing great science.

I also want to thank my committee members – Prof. David Tirrell, Prof. Barbara Wold and Prof. Michael Elowitz – for their time providing me with constructive feedback on my research and dissertation. Working with them and the students/post-docs from their labs (Kai, Graham and Peng to name just a few), I was able to rapidly get up to speed on high-throughput methods that became core parts of my thesis.

At Caltech, I am fortunate to have met many great colleagues and friends who shared my passion for science. At the Kornfield lab, I am lucky to be mentored by Amy on biology techniques that I use to this day. I'm also grateful to be able to work with Dan, Karthik, Tiziana, and others in the lab who helped me move my projects forward. I also can't forgot the wonderful core facilities we have on campus. To Igor and Vijaya at the Caltech Genomics Laboratory, thank you for helping me generate gigabyte after gigabytes of sequencing data. To the amazing folks (Michael, Annie, and others) at the Proteomics Exploration Lab, thank you for teaching me how to run proteomics experiments and for troubleshooting my

experiments again and again. I'm also grateful to have met many great friends through some of Caltech's amazing student clubs: Peng, Yapeng, Xiaozhe, Yandong, Ronghui, Andy, Seth, Phillip, and countless others, thank you for your friendship throughout the years.

I'm also fortunate to be part of the Caltech Biotechnology Leadership Program (BLP), spearheaded by our newest Nobel Laureate Prof. Frances Arnold and managed by Dr. Kimberly Mayer. One of the highlights and defining moments of my graduate study was to be able to join an industry internship in my field of study. The program enabled me to take on this challenge, which helped me explore my career interests and broaden my view of scientific research. Also, to my fellow BLP trainees (Josh, Pradeep, Bradley, Kevin, and Kelly, in particular), you guys made all of this possible and I benefited tremendously learning from you.

My projects brought me in touch with many collaborators around the world and I want to thank each of them for their contributions. First, Dr. Peter Campochiaro's lab has been instrumental in supporting my exploration of drug effect in mouse models. Also, Dr. Maria Cristina Kenney and Dr. Hugo Quiroz's labs for opening new chapters in the drug research project. I also want to thank our Italian colleagues, Carmela and Fausta, who helped me get started on the polymer biocompatibility project. Lastly, I want to thank the Allegro team members for their support throughout the years. To Dr. John Park in particular, thank you for helping me on the drug research projects and for your years of scientific support.

Lastly, I want to thank my friends and family who have always been there for me. To Michael, Anthony, Sam, Aranda, Martina, and Alex, your friendships means the world to me, so thank you from the bottom of my heart. To my family, mom and dad, thank you for raising me to be an inquisitive student of nature and for all of the sacrifices you made to make my journey possible. Finally, I want to thank my wife, Justine, for being there for me through all of the ups and downs of graduate school. I honestly don't know if I would be able to do it without you and thank you for constantly reminding me that there is a life outside of work!

Abstract

Diseases of the retina affect hundreds of millions of patients worldwide, with limited treatment options available. ALG-1001 is an investigational drug that showed success in mitigating disease symptoms in animal models and improved patient vision in multiple clinical trials. To gain a better understanding of the drug's mechanism of action, RNA sequencing (RNA-seq) and shotgun proteomics were employed to study the drug-induced transcriptome change in retinal tissue and cell culture models. **Chapter 2** focuses on application of this approach in an animal model of the disease that showed the drug can reversely modulate hypoxia-activated angiogenesis and inflammation gene expression changes. **Chapter 3** discusses the study of drug-induced transcriptome response in two cell culture models relevant to pathophysiology of the retinal diseases. **Chapter 4** explores retinal cell transcriptome after short and long-term exposure to disease-relevant hypoxia condition and after hypoxia recovery. **Appendix A** documents our shotgun proteomics protocol and includes results from the application of this method in the study of drug mechanism.

Typical RNA-seq studies use few biological replicates for differential expression analysis, mainly due to the high cost of generating sequencing data. As a result, not all comparisons have the proper statistical power, which result in false positives and false negatives that can lead the researcher to the wrong conclusion. **Chapter 5** discusses a novel algorithm and software that help users perform quality control of their dataset to identify whether the appropriate sample size was used for differential gene discovery. The chapter covers demonstration of the software with four publicly available RNA-seq datasets to illustrate its utility.

Bioresorbable vascular scaffolds (BVSs) are the application of biocompatible polymer in the treatment of coronary heart disease, one of the leading causes of death worldwide. BVSs are designed to replace metal stents, which stay permanently in the body after surgery and can lead to various complications, such as lethal thrombosis. In contrast, BVSs provide the necessary support and are resorbed by the body to leave behind a healthy artery after 2-3 years. Improving on the existing BVS material, **chapter 6** explores a new polymer nanocomposite that increases the structure's radial strength in a thinner profile and provides radio-opacity to enhance surgery success.

Published Content and Contributions

Shao Z and Kornfield J. “ERSSA: Empirical RNA-seq Sample Size Analysis.” *R package version 0.99.7* (2018). DOI: 10.18129/B9.bioc.ERSSA.

Shao Z. conceived the project and wrote the software package.

Table of Contents

List of Figures

Figure *Page*

List of Tables

Table *Page*

Chapter 1. Introduction

The human circulatory system is composed of a complex network of blood vessels that transport nutrients and oxygen to tissues and carry away metabolic waste. Disruption of this life-sustaining system can lead to tissue damage, disease, and even death. In many chronic diseases, abnormal growth of new blood vessels (called angiogenesis) sustains and aggravates the condition[1]. In other disorders, deterioration of existing vasculature can lead to irregular blood supply resulting in tissue necrosis[2]. The first three sections of this chapter (**Chapter 1.1-1.3**) discuss the vascular diseases of the retina, development of an investigational drug that targets these disorders and application of high-throughput methods to study the effect of the drug. The next section (**Chapter 1.4**) introduces the development of an analytical framework and an accompanying software that improves existing high-throughput transcriptome data analysis. The last section (**Chapter 1.5**) explores characterization of a nanocomposite biomaterial that has the potential to advance treatment of coronary heart disease.

1.1 Pathology and treatment of retinal diseases

The human eye is composed of a diverse variety of cells that are organized in a complex three-dimensional architecture to enable a central sensory function: visual perception of the world around us[3,4]. In the eye, light first enters through the cornea and is focused by the lens onto the retina. Within the retina, photoreceptors (cones and rods) detect and convert the light to chemical signals, which are passed to retinal neurons and ultimately to the brain. Contrary

to our intuition, over most of the retina, photons first pass through several layers of neurons and blood vessels before reaching the photoreceptors. These additional cell layers disperse and absorb the light, reducing spatial resolution. In contrast, the macula is an approximately 6mm dimeter region of the retina, where the neurons and the anterior blood vessels are mostly absent, thus allowing the light to more directly hit the photoreceptor cells. The macula also contains the highest density of cone photoreceptors, which provide us our high-resolution colored vision. The combination of high cell density and relatively lack of capillaries gives this highly metabolically-active tissue a predisposition for degeneration and disease[5,6].

In developed nations, degenerative diseases of the retina represent the leading cause of adult blindness[7]. Age-related Macular Degeneration (AMD) and Diabetic Retinopathy/Diabetic Macular Edema (DR/DME) are two major disorders that predominantly affect the macula. Based on the most recent statistics, AMD affects 11 million Americans (170 million worldwide), and accounts for more than half of all blindness and visual impairment in the industrialized countries[8,9]. Similarly, DR/DME affect approximately 4 million Americans (93 million worldwide), and is the largest cause of blindness under the age of 55 in the developed world[10,11].

Many factors, including both genetic predisposition and environmental stress, contribute to the onset of these diseases. Over time, presence of stress factors lead to development of chronic hypoxia and inflammation[12], which contribute to develop of pathological angiogenesis (pathophysiology of DR/DME in **Figure 1.4** and AMD in **Figure 1.5**)[7,13,14]. In early stages of AMD, the retina is inflamed with activated microglia cells (resident

macrophages of central nervous system) and show sign of degenerative blood vessel loss[15]. As AMD progresses, the retina become plagued by accumulation of extracellular debris in the posterior region in deposits called drusen[16]. Drusen formations hinder diffusion between the vasculature and retina, and their presence attracts inflammatory macrophages to the area[17,18]. The combination of these processes drive chronic hypoxia and inflammation in the AMD retina, contributing to the disease progression. In DR/DME, hyperglycemia is associated with damage and regression of retinal vasculature, creating a hypoxic/ischemic micro-environment in the tissue[19,20]. High blood sugar and excess lipid conditions also promote release of inflammatory mediators by the neural retina, leading to leukocyte recruitment. In both diseases, hypoxic and inflammatory conditions converge to drive cellular release of a variety of pro-angiogenic factors, including Vascular Endothelial Growth Factor (VEGF), a potent angiogenesis activator[21]. These signaling then lead to formation of immature vessels that are often poorly-developed and leak fluid, resulting in edema, bleeding and cell death in the retina/macula[13]. These degenerative conditions progressively get worse, causing vision impairment and eventually blindness.

Currently, there are two main treatment options for the described retinal diseases. Before the introduction of anti-VEGF pathway antibodies, laser therapy was the most commonly used approach. One type of laser therapy burns the peripheral retinal tissue, while sparing the macula to reduce the metabolic demand and alleviate the hypoxia/ischemia condition[20]. A second type of laser therapy targets specific leaky blood vessels to seal the leakage and reduce edema/bleeding[22]. However, laser-based therapies generally damage the peripheral

vision and do not usually lead to vision improvement (they stabilize vision deterioration). Furthermore, neovascularization often re-emerges in many of the patients after surgery, making this treatment option only a short-term solution[23]. The current standard of care uses anti-VEGF antibodies (Ranibizumab, Aflibercept and Bevacizumab[15,24]) that specifically bind with VEGF and disrupt the activation of its signaling pathway to suppress retinal neovascularization. While significantly more successful than laser therapy, the antibody treatments are: a) expensive[25], b) require burdensome intraocular (into the eye) injection every 30-60 days[26], c) is ineffective in more than 25-50% of the patients[27,28]. Due to these drawbacks, an alternative and more efficacious therapy is needed to address the limitations of the existing standard of care.

1.2 Discovery of an oligopeptide therapeutic agent – ALG-1001

An investigational oligopeptide drug named ALG-1001 (**Figure 1.1**) was invented by Allegro Ophthalmics, LLC. and is currently undergoing clinical trial in the described retinal diseases. In an Phase I clinical trial, 15 subjects with very advanced forms of DME were recruited to receive three monthly intraocular injections, followed by three months of additional monitoring[29]. At the end of the study, no major drug-related side effects were observed in all fifteen patients and more than half of the subjects (8 out of 15 patients, including several that are refractory to existing antibody therapy) demonstrated a significant improve in vision (3 BCVA lines of improvement in visual acuity, **Figure 1.2**). Since this initial study, additional phase I/II trials with DR, DME and AMD patient cohorts were

completed and continued to demonstrate ALG-1001 is well-tolerated and improved vision with no drug-related side effects reported to date[30–32]. Based on these positive results, the company recently announced the plan to enter the drug into a phase III trial in DME, slated to start in early 2019[33].

ALG-1001 was invented to structurally mimic a class of proteins that contain a common arginine, glycine, aspartic acid (RGD) amino acid sequence. The RGD sequence, in context of its flanking amino acids, has been studied extensively in the past for its natural role as the key recognition sequence of many important extracellular matrix proteins[34]. RGD containing peptides and their derivatives have also been investigated for potential therapeutic value in applications including tumor therapy, tissue engineering, and drug delivery[35–37]. The RGD motif was first isolated in 1984 by Pierschbacher and Ruoslahti as the specificity determining sequence in fibronectin proteins[38]. Later research found that the motif facilitates the normal binding of several important groups of extracellular matrix proteins to a class of cell receptors called integrins[39]. Today, 24 integrin heterodimers have been discovered in mammals, and their expression pattern on various cell types facilitate both normal and pathogenic cellular physiology. In particular, elevated expressions of certain integrins (*e.g.* αvβ3, αvβ5 and α5β1) are often observed on developing blood vessels during pathological angiogenesis[40].

In ocular tissues, it was previously reported that AMD patients express high level of αvβ3 integrins while both αvβ3 and αvβ5 integrins were elevated in DR patients[41]. Similarly, α5β1 elevation has been observed in neovascularization of the ocular tissue in animal models[42]. Since these integrins are selectively up-regulated in newly-formed blood vessels, it was

hypothesized that their expression is important for angiogenesis. Indeed, using the chick Chorioallantoic Membrane (CAM) model of angiogenesis, Brooks and Drake found that anti-integrin antibodies were able to inhibit vessel growth[43,44]. Around the same time, various research groups hypothesized that short RGD-containing peptides may also specifically bind with integrins and disrupt their normal involvement in angiogenesis. In a 1996 study, mouse subcutaneous injection of a cyclic RGD peptide was shown to reduce vessel proliferation by 76% in the mouse retina[45]. In the same year, another independent study by Friedlander and coworkers reached the same conclusion using the same peptide and a different mouse model[41]. These results and additional confirmatory studies demonstrate that RGD-containing peptides can strongly inhibit angiogenesis through disruption of pro-angiogenic integrin function.

In light of these early research, ALG-1001 was invented as an RGD mimic for treatment of neovascularization in the eye. Based on the RGD sequence, Dr. John Park (Allegro Ophthalmics, LLC.) replaced the carboxyl group on the aspartic acid with a sulfonate group, *i.e.* cysteic acid. In drug research, sulfonate and carboxylate groups are classical example of structurally distinct, yet bio-functionally equivalent molecular fragments[46]. In the original hypothesis, the placement of sulfonate group was proposed to induce a stronger interaction with the integrin binding pocket and thus increases the drug's potency. Although more recent biochemical experiments call this hypothesis into question, concurrent studies revealed that ALG-1001 is a strong inhibitor of angiogenesis in animal models.

Using three different mouse models that mimic human retinal diseases, Dr. Peter Campochiaro (Johns Hopkins University) and his team demonstrated that ALG-1001 is effective in reducing ocular neovascularization (personal communication). In the Laser-induced Choroidal Neovascularization (CNV) mouse model[47], ALG-1001 injection reduced neovascularization by 43% compared to vehicle control. In another model named Oxygen-induced Retinopathy (OIR)[48], ALG-1001 was shown to reduce vessel growth by up to 54% five days after injection. In a third study of a mouse model that over-expresses human VEGF and subsequently develops retinal angiogenesis, ALG-1001 and the current standard of care, Ranibizumab, were tested in a combination study. In this experiment, both drugs comparably reduced subretinal neovascularization, but there was a further reduction when the two drugs were used in combination. The synergistic effect suggests ALG-1001 may act through a molecular pathway that differs from existing anti-VEGF treatment (as the effects appear to be additive). Collectively, these *in vivo* studies show that ALG-1001 is effective in modulating retinal neovascularization.

1.3 ALG-1001 mechanism of action study

Although ALG-1001 was designed to be a more potent inhibitor of integrin than RGD, recent studies found the drug's binding affinity with various integrins are, in fact, weak. In one study at the Swenson lab (USC), Fluorescence Polarization Assay found micro-molar binding constants between ALG-1001 and αvβ3 (K_d=0.9uM), α5β1 (K_d>1uM) and αvβ5 (K_d>1uM) integrins, in contrast to positive control between vitronectin and αvβ3 integrin (K_d

≅ 8nM) (personal communication). Another study performed at the Kornfield lab (Caltech) used Surface Plasmon Resonance (SPR) and reached the same conclusion (K_d ≅ 2mM for ALG-1001 with α5β1, while a control RGD peptide has a small K_d ≅ 12nM, unpublished). Altogether, these experiments suggest the binding affinity between ALG-1001 and recombinant integrins are weak and non-specific. However, these studies were also done with recombinant integrins that do not perfectly mimic the endogenous structure of integrins[49] (*e.g.,* the flexible and condition-dependent conformation of the extracellular domain of integrin heterodimers[50]).

Based on these experiments, the hypothesis that ALG-1001 exerts its anti-angiogenesis effect through binding with pro-angiogenic integrins has not been confirmed and its precise mechanism of action requires further investigation. Toward the goal of uncovering the drug's mechanism of action, my main thesis project aims to discover the biological pathways that confer the drug's therapeutic effect. To do this, I evaluated two main methods to quantify the drug-induced changes: 1) in the transcriptome using RNA sequencing (RNA-seq) on polyadenylated messenger RNAs (mRNAs), and 2) in the proteome using mass-spectrometry-based shotgun proteomics. The two methods naturally complement each other, as evidence in both the transcriptome and the proteome can reveal the drug-induced biological regulation. Additionally, the two approaches have their unique advantages and drawbacks that warrant their combined application.

Methods measuring the proteome are powerful, as they directly study the expression level of cellular proteins, which perform a majority of functions in the cell[51]. Shotgun proteomics

started to become popular in the early 2000s as a high-throughput method to probe the proteome[52]. Despite its high-throughput nature, shotgun proteomics rely on mass spectrometry for measurement and current methods have little to no detection power for medium to low expressing proteins, on a practical scale[53,54]. On the other hand, nucleic acid-based methods that probe the transcriptome can overcome this limitation through polymerase-based signal amplification[55]. As a result, methods such as RNA-seq have a much larger dynamic range of detection compared to typical shotgun proteomics and enables identification of changes in even the low expressing genes[56]. However, it has previously been reported that only around 40% of protein expression levels can be explained by mRNA dynamics while the rest are due to protein translation and degradation regulatory mechanisms[57]. Although recent studies suggest stronger correlations between the proteome and transcriptome (possibly due to better normalization methods and experimental protocols)[58,59], the importance of protein expression regulation warrants observation at the proteome level. Faced with an unknown mechanism of action, I worked with Dr. Kai Yuet and Dr. Graham Hamblin in the Tirrell group (Caltech) to evaluate shotgun proteomics as an unbiased method to probe the effects of ALG-1001 at the level of the proteome.

In shotgun proteomics, proteins are detected using liquid chromatography and tandem mass spectrometry (LC-MSMS) and analyzed to identify pattern of molecular fragments that match known protein sequences[53,60]. To detect differences in transient protein expression in response to a drug, it is necessary to differentiate between proteins synthesized before and after treatment, and between treated and control samples. To do this, Pulse Stable Isotope

Labeling by Amino Acids in Cell Culture (pSILAC) is used to provide relative protein ratio in two samples (one of each condition) by incorporating different isotopically-labelled amino acids (Lysine and Arginine) into newly-synthesized proteins (**Figure 1.3**)[61]. To obtain sufficient signal-to-noise ratio for the isotopically-labelled proteins, it is beneficial to suppress the large background of unlabeled proteins using a technique pioneered in the Tirrell lab at Caltech: Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT), which introduces the noncanonical amino acid azidohomoalanine (AHA) in place of methionine[62] to enable covalent capture of nascent proteins. When AHA is introduced at the same time as the isotopically-labeled amino acids, BONCAT enables the enrichment of the isotopically-labeled proteins[63] and has been successfully applied in recent years in many biological models including mammalian cell lines[63–65], mouse *ex vivo* brain tissue[66] and *Caenorhabditis elegans*[67]. In our study, removal of background protein using BONCAT is important due to a relatively short drug treatment period (<48 hours) versus the average protein half-life (*e.g.*, estimated to be around 100 hours in primary cells[68]).

To validate our technique, we performed a preliminary BONCAT-pSILAC experiment comparing a RGD-containing peptide versus control peptide, which successfully identified several hundred differentially expressed (DE) proteins in mammalian cell culture. Many of the regulated proteins were found to function in cell to cell adhesion (**Appendix A.2**), in agreement with RGD disruption of integrin-mediated cellular interaction. However, when we applied BONCAT-pSILAC to search for proteomic changes in response to ALG-1001, we did not identify any DE proteins (**Appendix A.3**). Subsequent RNA-seq experiments

reveal the measurable effect of ALG-1001 *in vitro* is weak (**Chapter 3**), even though stronger changes in the transcriptome are readily measured *in vivo* in the OIR mouse model (**Chapter 2**). Although it is beyond the scope of this thesis, proteomic effects of ALG-1001 that are below the sensitivity of BONCAT-pSILAC *in vitro* may be measurable in a disease model *in vivo*.

In contrast to the limited detection power of shotgun proteomics, RNA-seq is capable of measuring even low-expressing genes. In our hands, BONCAT-pSILAC detected only a few thousand highly expressed proteins, which limit the search for DE proteins to a small subset of all proteins expressed. The ability to routinely detect more than ten thousand genes using RNA-seq sheds new light on the effect of ALG-1001 in various biological models.

RNA-seq, a high-throughput method of probing mRNA expression, was first applied to study the complex eukaryote transcriptome by the Wold lab at Caltech[69]. Since then, RNA-seq rapidly became the standard technique to study cellular transcriptome, largely replacing previously popular methods such as microarray and SAGE[70]. In a typical RNA-seq experiment, mRNA-derived complementary DNAs (cDNAs) are sequenced concurrently to generate millions of reads that contain sequence information[71]. The reads can then be mapped to the reference genome/transcriptome to tally the number of reads that map onto each gene in the genome. The tabulation of the number of mapped reads per gene is repeated for each sample in the experiment and statistical methods are then applied to identify genes that are DE between the groups[72,73]. Finally, the list of DE genes, together with corresponding fold change and measure of statistical significance, serve as inputs for functional analysis to

search for regulatory pathways, gene ontology groups and transcription factors that are correlated with the observed expression variation.

Adoption of RNA-seq as the *de facto* tool for study of the transcriptome has been rapid. Since its introduction in 2008, over 14,000 RNA-seq-related papers have been published (PubMed, late 2018), including more than 3,000 in 2017 alone. There are several reasons to RNA-seq's popularity. First, the technique has a high dynamic range (tens of thousands of genes across four or more orders of magnitude in expression level can be detected simultaneously)[56]. Second, RNA-seq typically has high reproducibility, where it is not uncommon to see technical replicates with $R^2 > 0.99$. Third, sample preparation and sequencing is increasingly streamlined as a rapid increase in the number of users spurred development of commercial products. Fourth, the method can be easily adapted to study any organism, a major advantage over microarrays. Finally, there exists an active bioinformatics community (such as the Bioconductor project[74]) that develops analytical tools to support computational analysis of RNA-seq data. As a result, biomedical application of RNA-seq has become generally accepted, particularly in fields such as cancer research, disease diagnostics and drug discovery[75].

In our research, we applied comparative RNA-seq to identify the transcriptome regulation modulated by hypoxia and ALG-1001 in various biological models: *in vivo* OIR mouse retinal neovascularization model (**Chapter 2**), *in vitro* human endothelial and immune cell culture models (**Chapter 3**), and retinal cells conditioned with short to long-term hypoxia

and hypoxia recovery (**Chapter 4**). We also documented our BONCAT-pSILAC protocol and the experimental results in **Appendix A**.

1.4 Sample size in comparative RNA-seq experiments

It is well established that comparative RNA-seq experiments require biological replicates for statistical testing[76]. Increasing the number of biological replicates improves both statistical power and accuracy of gene expression estimation[72,77,78]. However, RNA-seq experiments are expensive (currently, each sample typically cost several hundred dollars in library preparation and sequencing alone) and sample procurement may be challenging in certain experimental settings. As a result, RNA-seq studies often include as few as two or three biological replicates per experimental condition, which put the results of statistical test in question.

Selecting the appropriate number of replicates (sample size) for each condition examined in a specific RNA-seq experiment is challenging, as several determining factors are generally not known *a priori*[79]. First, when the true inter-condition difference (effect size) is small, more replicates are needed for differential expression discovery. Second, when biological variance is high among samples of the same condition, more replicates are needed to improve statistical power. Although several software packages have been released in recent years to help researchers choose the appropriate sample size for RNA-seq studies, Poplawski and Binder (2017) found that all of the tools failed to accurately estimate sample size even for substantial changes in gene expression (fold change $\geq$ 1.5)[80]. This systematic study

performed simulation-based assessment on six RNA-seq sample size calculation tools that support sample size calculation with multiple genes in consideration. The results from the evaluation showed that the six tools provided widely different results that are highly dependent on the effect size in consideration and none of them gave satisfactory agreement with the "ground truth" (which is known for the simulated data set used in the study). At present, RNA-seq sample size selection remains challenging and additional tools are needed to ensure accurate differential expression discovery.

In **Chapter 5**, we describe a solution to this problem with an algorithm that does not rely on data simulation or *a priori* assumed parameters. Instead, we approach the problem by analyzing an existing data set that a user has already acquired to evaluate the benefit they can expect to achieve with additional replicates. The method, called Empirical RNA-seq Sample Size Analysis (ERSSA), uses a subsampling technique to compute the effect of increasing number of replicates on the number of DE genes identified. Using the trend in the amount of marginal improvement as sample size increases, the analytical framework allows the user to evaluate the potential benefit of procuring additional biological replicates. The algorithm has been implemented in the statistical programming language R and is publicly available on Bioconductor at http://bioconductor.org/packages/ERSSA/.

1.5 Application of Bioresorbable Vascular Scaffolds in coronary heart disease

Coronary heart disease (CHD) is a leading cause of mortality in the world. In the US alone, CHD is associated with one in every five deaths each year[2]. In CHD, vessel-occluding lipid

deposits build up in the coronary artery and obstructs or complete blocks blood flow to the heart muscle, eventually leading to angina and heart attack[81]. In recent decades, the main therapy for CHD is to insert a permanent metal stent at the site of occlusion to open and support the artery and to maintain blood flow[82]. While metal stents provide tremendous therapeutic value, they also have some important side effects, including a higher risk of life-threatening late thrombosis[83]. With more than 1 million stents implanted each year in the US alone as of 2008, there is an urgent need to develop safer therapies for CHD[84].

Bioresorbable vascular scaffold (BVS), such as Abbott Vascular's FDA-approved Absorb®, recently emerged as a promising alternative to metal stents[85]. Absorb® is composed of biodegradable poly L-lactide (PLLA) that supports the artery for the first few months and then completely degrades in 2-3 years[86]. A five year follow-up clinical trial involving 101 patients showed that artery regains the ability to dilate and pulsate, with no incidence of late thrombosis in patients[87]. Despite this initial success, a larger follow-up study found an increase in thrombosis associated with BVS (~1.3% prevalence rate) relative to metal stents (~0.6%)[88]. Two potential explanations provided by clinicians include: 1) BVSs contains thicker struts compared to metal stents, and 2) under-deployment of BVSs over fear of scaffold fracture and inability to image the deployed diameter during surgery. These concerns motivate development of a BVS device that has thinner strut as well as superior radial strength to reduce the chance of thrombosis.

Working with a team of collaborators in the Kornfield lab and at ENEA Portici in Italy, we explored the application of Tungsten Disulfide nanotubes (WSNT) as a reinforcing agent

that could increase the strength and x-ray scattering contrast of the BVS. My collaborators characterized the effects of WSNT on the processing behavior and semi-crystalline morphology of PLLA-WSNT nanocomposites. I characterized the biocompatibility of this reinforcing agent and of the nanocomposite. Our combined results are included in **Chapter 6**.

1.6 Figures and Tables

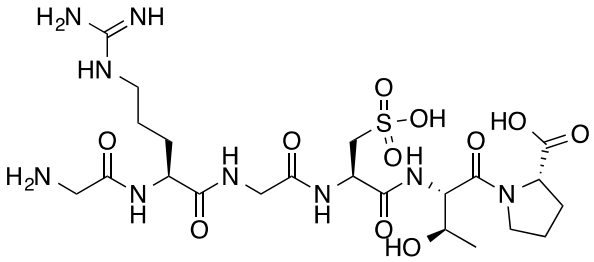

Figure 1.1 Molecular structure of ALG-1001 (Gly-Arg-Gly-Cys(acid)-Thr-Pro)
ALG-1001 is structurally designed to mimic RGD-containing peptides. Aspartic acid in the RGD sequence is replaced by a noncanonical amino acid – cystic acid[89].

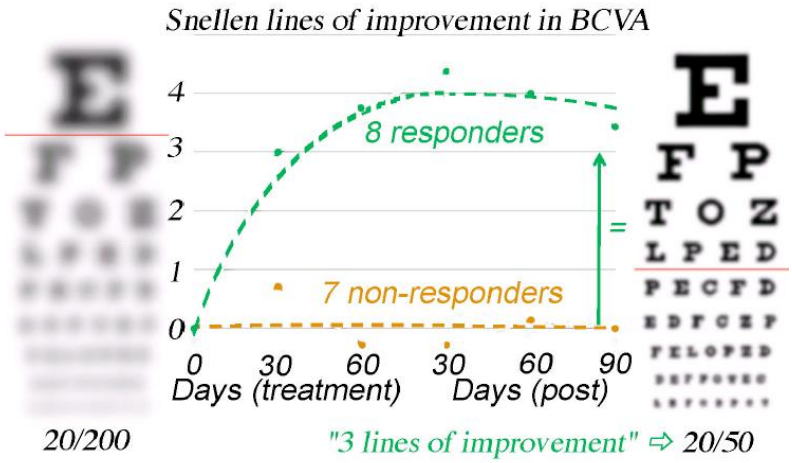

Figure 1.2 ALG-1001 Phase I clinical trial in DME

Phase I study of 15 "end-stage" DME patients with severe vision impairment. Three monthly ALG-1001 intraocular injections were performed follow by three months of follow-ups. No other treatment arms were included in this trial. At the end of 150 day trial, 8 out of 15 patients showed significant improvement in a visual acuity test (green dots), while 7 out of 15 patients showed no further visual deterioration (yellow dots). Dashed lines outline the trend in vision change in the two responder groups. Vertical, solid green arrow indicates the magnitude of improvement illustrated by the Snellen chart simulation to the left and right of the graph.

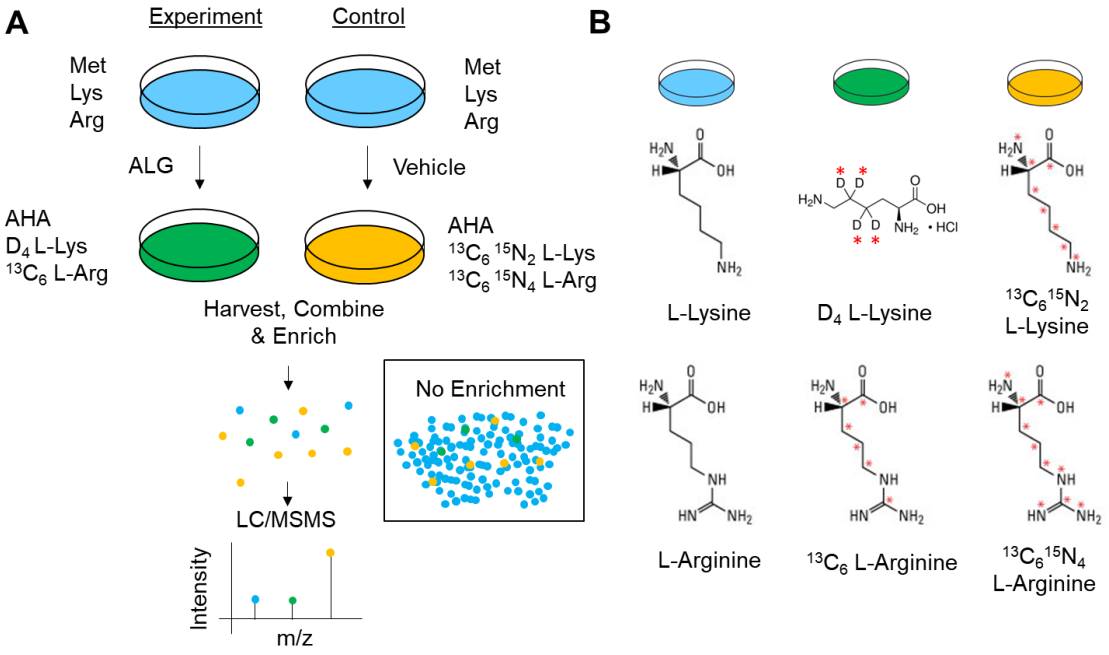

Figure 1.3 BONCAT with pSILAC schematic

Illustration describes the major steps in a BONCAT with pSILAC experiment. Adapted from a review publication by Yuet and Tirrell[90]. **A.** In a typical BONCAT with pSILAC experiment, cells in two groups (treated "experiment" and untreated "control") are grown in conventional media (blue) to the desired confluency. At the onset of treatment, the conventional media is removed and replaced with media containing 30 times more AHA than methionine, and all of the lysine and arginine replaced by isotopically labelled counterparts. In the depicted study, the new media for the "experiment" also contains a peptide of interest (ALG denotes ALG-1001 in the green medium) while "control" does not. After the treatment period, cells are harvested and lysed, proteins combined in 1:1 ratio and enriched through click-reaction chemistry (unenriched protein mixture shown in box labeled "no enrichment"). The enriched proteins are then characterized using liquid chromatography tandem-mass spectrometry (LC-MSMS), giving the "triplets" of peaks associated with background protein (blue), newly-synthesized protein from the "experiment" (green) and newly-synthesized protein from the "control" (yellow). Met is methionine, Lys is lysine, Arg is arginine and m/z is mass to charge ratio. **B.** Conventional lysine and arginine (*left*) are compared with "heavy" counterparts (*right*) and "medium" counterparts (*center*), where asterisks indicate sites where deuterium (D) replaces hydrogen, ^{13}C replaces ^{12}C and/or ^{15}N replaces ^{14}N.

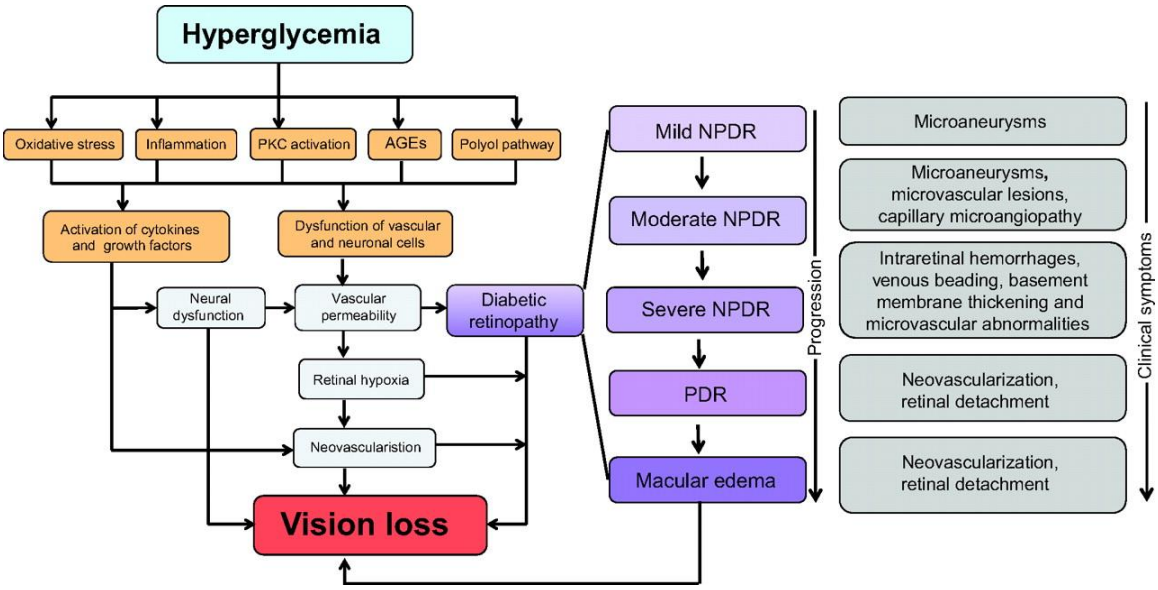

Figure 1.4 Model of diabetic retinopathy
Diabetes-related hyperglycemia induce generation of oxidative stress, inflammation, advanced glycation end products (AGEs) and activation of Protein Kinase C (PKC) and polyol pathways. These activated factors/pathways contribute to hypoxia and neovascularization in the eye, leading to vision loss. NPDR is non-proliferative diabetic retinopathy, PDR is proliferative diabetic retinopathy. Figure obtained from [91] with sharing allowed under a CC-BY license.

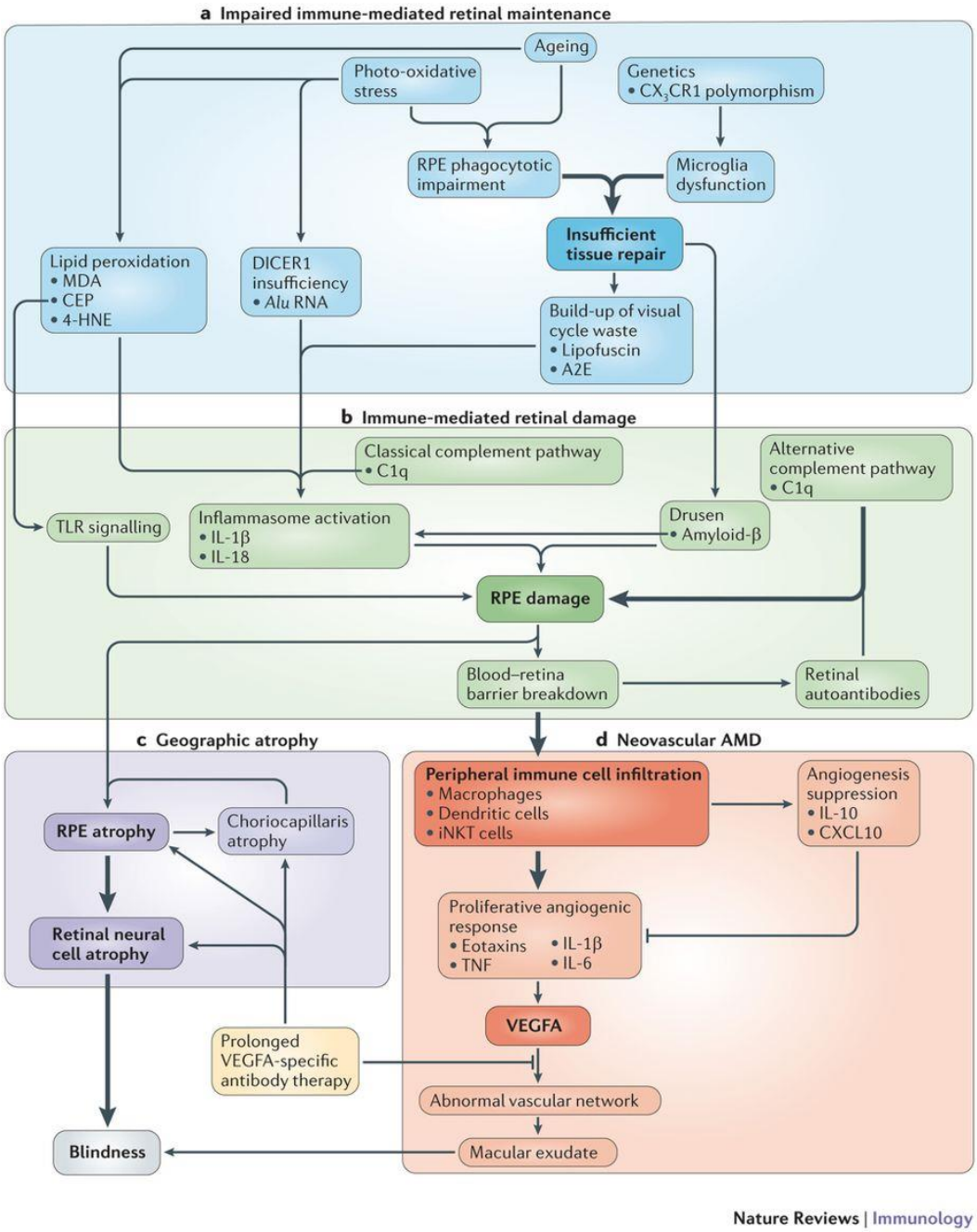

Figure 1.5 Model of macular degeneration

In AMD, both environmental and genetic factors activates inflammation and cause RPE damage. An accumulation of activated immune cells and dysregulated RPE cells then contribute to tissue hypoxia, which activates pathological angiogenesis that can eventually lead to blindness. Reprinted with permission from Ambati *et al.*[92].

1.7 References

1. Folkman, J. Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nat. Med.* **1**, 27–30 (1995).
2. Chilton, R. J. Pathophysiology of Coronary Heart Disease: A Brief Review. *J. Am. Osteopath. Assoc.* **104**, 5S-8S (2004).
3. Oyster, C. W. *The Human Eye: Structure and Function*. (Sinauer Associates, 1999).
4. Riordan-Eva, P. & Augsburger, J. J. *Vaughan & Asbury's General Ophthalmology, 19th Edition*. (McGraw-Hill Education / Medical, 2017).
5. Antonetti, D. A. *et al.* Diabetic Retinopathy. *Diabetes* **55**, 2401–2411 (2006).
6. Newman, A. M. *et al.* Systems-level analysis of age-related macular degeneration reveals global biomarkers and phenotype-specific functional networks. *Genome Med.* **4**, 16 (2012).
7. Radeke, M. J., Peterson, K. E., Johnson, L. V. & Anderson, D. H. Disease susceptibility of the human macula: Differential gene transcription in the retinal pigmented epithelium/choroid. *Exp. Eye Res.* **85**, 366–380 (2007).
8. Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M. & Bhattacharya, S. S. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.* **11**, 273–284 (2010).
9. Pennington, K. L. & DeAngelis, M. M. Epidemiology of age-related macular degeneration (AMD): associations with cardiovascular disease phenotypes and lipid factors. *Eye Vis.* **3**, (2016).
10. Leading Causes of Blindness in the U.S. Available at: http://www.nei.nih.gov/health/fact_sheet.asp.
11. Holekamp, N. M. Overview of diabetic macular edema. *Am. J. Manag. Care* **22**, s284–s291 (2016).
12. Whitcup, S. M., Nussenblatt, R. B., Lightman, S. L. & Hollander, D. A. Inflammation in Retinal Disease. *Int. J. Inflamm.* (2013). doi:10.1155/2013/724648
13. Friedlander, M. Fibrosis and diseases of the eye. *J. Clin. Invest.* **117**, 576–586 (2007).
14. Tang, J. & Kern, T. S. Inflammation in Diabetic Retinopathy. *Prog. Retin. Eye Res.* **30**, 343–358 (2011).
15. van Lookeren Campagne, M., LeCouter, J., Yaspan, B. L. & Ye, W. Mechanisms of age-related macular degeneration and therapeutic opportunities. *J. Pathol.* **232**, 151–164 (2014).
16. Lim, L. S., Mitchell, P., Seddon, J. M., Holz, F. G. & Wong, T. Y. Age-related macular degeneration. *The Lancet* **379**, 1728–1738 (2012).
17. An, E. *et al.* Secreted Proteome Profiling in Human RPE Cell Cultures Derived from Donors with Age Related Macular Degeneration and Age Matched Healthy Donors. *J. Proteome Res.* **5**, 2599–2610 (2006).
18. An, E., Gordish-Dressman, H. & Hathout, Y. Effect of TNF-? on human ARPE-19-secreted proteins. *Mol. Vis.* **14**, 2292–2303 (2008).

19. Stitt, A. W., Lois, N., Medina, R. J., Adamson, P. & Curtis, T. M. Advances in our understanding of diabetic retinopathy. *Clin. Sci.* **125**, 1–17 (2013).
20. Cheung, N., Mitchell, P. & Wong, T. Y. Diabetic retinopathy. *The Lancet* **376**, 124–136 (2010).
21. Costa, C., Incio, J. & Soares, R. Angiogenesis and chronic inflammation: cause or consequence? *Angiogenesis* **10**, 149–166 (2007).
22. Talwar, D. *et al.* Contrast sensitivity following focal laser photocoagulation in clinically significant macular oedema due to diabetic retinopathy. *Clin. Experiment. Ophthalmol.* **29**, 17–21 (2001).
23. Macular Photocoagulation Study Group. Persistent and recurrent neovascularization after laser photocoagulation for subfoveal choroidal neovascularization of age-related macular degeneration. Macular Photocoagulation Study Group. *Arch. Ophthalmol. Chic. Ill 1960* **112**, 489–499 (1994).
24. Abcouwer, S. F. & Gardner, T. W. Diabetic retinopathy: loss of neuroretinal adaptation to the diabetic metabolic environment. *Ann. N. Y. Acad. Sci.* **1311**, 174–190 (2014).
25. Parikh, R. *et al.* A Multinational Comparison of Anti-Vascular Endothelial Growth Factor Use: The United States, the United Kingdom, and Asia-Pacific. *Ophthalmol. Retina* **3**, 16–26 (2019).
26. Hussain, R. & Ciulla, C. Addressing the Anti-VEGF Treatment Burden. Available at: https://www.reviewofophthalmology.com/article/addressing-the-antivegf-treatment-burden. (Accessed: 19th May 2019)
27. Ehlken, C. *et al.* Switch of anti-VEGF agents is an option for nonresponders in the treatment of AMD. *Eye* **28**, 538–545 (2014).
28. Singer, M. A., Kermany, D. S., Waters, J., Jansen, M. E. & Tyler, L. Diabetic macular edema: it is more than just VEGF. *F1000Research* **5**, (2016).
29. Allegro Ophthalmics Receives Fda Ind Approval For Phase Ii Clinical Studies With First In Class Integrin Peptide Therapy Alg-1001 – Allegro Ophthalmics – From Theory to Therapy.
30. A Clinical Trial Designed to Evaluate the Safety and Efficacy Study of Luminate® in Inducing PVD (Posterior Vitreous Detachment) in Non-Proliferative Diabetic Retinopathy - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02435862. (Accessed: 11th September 2018)
31. A Safety And Efficacy Study Of Alg-1001 In Human Subjects With Symptomatic Focal Vitreomacular Adhesion - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02153476. (Accessed: 11th September 2018)
32. A Safety And Efficacy Study Of ALG-1001 In Human Subjects With Wet Age-Related Macular Degeneration - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT01749891. (Accessed: 11th September 2018)
33. Allegro Ophthalmics Prepares To Take Lead Compound Risuteganib To Phase 3, Appoints New CEO – Allegro Ophthalmics – From Theory to Therapy.
34. Ruoslahti, E. Rgd and Other Recognition Sequences for Integrins. *Annu. Rev. Cell Dev. Biol.* **12**, 697–715 (1996).

35. Mas-Moruno, C., Rechenmacher, F. & Kessler, H. Cilengitide: The First Anti-Angiogenic Small Molecule Drug Candidate. Design, Synthesis and Clinical Evaluation. *Anticancer Agents Med. Chem.* **10**, 753–768 (2010).
36. Wang, F. *et al.* The Functions and Applications of RGD in Tumor Therapy and Tissue Engineering. *Int. J. Mol. Sci.* **14**, 13447–13462 (2013).
37. Park, J. *et al.* A review of RGD-functionalized nonviral gene delivery vectors for cancer therapy. *Cancer Gene Ther.* **19**, 741–748 (2012).
38. Pierschbacher, M. D. & Ruoslahti, E. Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Publ. Online 03 May 1984 Doi101038309030a0* **309**, 30–33 (1984).
39. Hynes, R. O. Integrins: Bidirectional, Allosteric Signaling Machines. *Cell* **110**, 673–687 (2002).
40. Ruegg, C. & Mariotti, A. Vascular integrins: pleiotropic adhesion and signaling molecules in vascular homeostasis and angiogenesis. *Cell. Mol. Life Sci.* **60**, 1135–1157 (2003).
41. Friedlander, M. *et al.* Involvement of integrins alpha v beta 3 and alpha v beta 5 in ocular neovascular diseases. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 9764–9769 (1996).
42. Umeda, N. *et al.* Suppression and Regression of Choroidal Neovascularization by Systemic Administration of an α5β1 Integrin Antagonist. *Mol. Pharmacol.* **69**, 1820–1828 (2006).
43. Brooks, P. C., Clark, R. A. F. & Cheresh, D. A. Requirement of vascular integrin alpha v beta 3 for angiogenesis. *Science* **264**, 569–571 (1994).
44. Drake, C. J., Cheresh, D. A. & Little, C. D. An antagonist of integrin alpha v beta 3 prevents maturation of blood vessels during embryonic neovascularization. *J. Cell Sci.* **108**, 2655–2661 (1995).
45. Hammes, H.-P., Brownlee, M., Jonczyk, A., Sutter, A. & Preissner, K. T. Subcutaneous injection of a cyclic peptide antagonist of vitronectin receptor–type integrins inhibits retinal neovascularization. *Nat. Med.* **2**, 529–533 (1996).
46. Nogrady, T. & Weaver, D. F. *Medicinal Chemistry: A Molecular and Biochemical Approach*. (Oxford University Press, USA, 2005).
47. Lambert, V. *et al.* Laser-induced choroidal neovascularization model to study age-related macular degeneration in mice. *Nat. Protoc.* **8**, 2197–2211 (2013).
48. Stahl, A. *et al.* The Mouse Retina as an Angiogenesis Model. *Invest. Ophthalmol. Vis. Sci.* **51**, 2813–2826 (2010).
49. He, Y., Wang, K. & Yan, N. The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein Cell* **5**, 658–672 (2014).
50. Ley, K., Rivera-Nieves, J., Sandborn, W. J. & Shattil, S. Integrin-based therapeutics: biological basis, clinical use and new drugs. *Nat. Rev. Drug Discov.* **15**, 173 (2016).
51. Cooper, G. M. The Molecular Composition of Cells. *Cell Mol. Approach 2nd Ed.* (2000).
52. Hernandez-Fernaud, J. R., Reid, S. E., Neilson, L. J. & Zanivan, S. Quantitative mass spectrometry-based proteomics in angiogenesis. *PROTEOMICS – Clin. Appl.* **7**, 464–476 (2013).

53. Steen, H. & Mann, M. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**, 699–711 (2004).
54. Zubarev, R. A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *PROTEOMICS* **13**, 723–726 (2013).
55. Ferreira, E. N. *et al.* Linear mRNA amplification approach for RNAseq from limited amount of RNA. *Gene* **564**, 220–227 (2015).
56. Nookaew, I. *et al.* A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. *Nucleic Acids Res.* **40**, 10084–10097 (2012).
57. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
58. Jovanovic, M. *et al.* Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038 (2015).
59. Li, J. J., Bickel, P. J. & Biggin, M. D. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**, e270 (2014).
60. Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
61. Mann, M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* **7**, 952–958 (2006).
62. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci.* **103**, 9482–9487 (2006).
63. Howden, A. J. M. *et al.* QuaNCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **10**, 343–346 (2013).
64. Bagert, J. D. *et al.* Quantitative, Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labeling by Amino Acids in Cell Culture. *Mol. Cell. Proteomics* **13**, 1352–1358 (2014).
65. Eichelbaum, K., Winter, M., Diaz, M. B., Herzig, S. & Krijgsveld, J. Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol.* **30**, 984–990 (2012).
66. Bowling, H. *et al.* BONLAC: A combinatorial proteomic technique to measure stimulus-induced translational profiles in brain slices. *Neuropharmacology* **100**, 76–89 (2016).
67. Yuet, K. P. *et al.* Cell-specific proteomic analysis in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2705–2710 (2015).
68. Mathieson, T. *et al.* Systematic analysis of protein turnover in primary cells. *Nat. Commun.* **9**, 689 (2018).
69. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
70. Xu, J. *et al.* Comprehensive Assessments of RNA-seq by the SEQC Consortium: FDA-Led Efforts Advance Precision Medicine. *Pharmaceutics* **8**, 8 (2016).

71. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
72. Seyednasrollah, F., Laiho, A. & Elo, L. L. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.* bbt086 (2013). doi:10.1093/bib/bbt086
73. Rapaport, F. *et al.* Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.* **14**, R95 (2013).
74. Bioconductor. Available at: https://www.bioconductor.org/. (Accessed: 14th September 2018)
75. Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D. & Craig, D. W. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* **advance online publication**, (2016).
76. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8**, 469–477 (2011).
77. Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8**, 1765–1786 (2013).
78. Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301–304 (2014).
79. Hoskins, S. P., Shyr, D. & Shyr, Y. Sample Size Calculation for Differential Expression Analysis of RNA-Seq Data. in *Frontiers of Biostatistical Methods and Applications in Clinical Oncology* 359–379 (Springer, Singapore, 2017). doi:10.1007/978-981-10-0126-0_22
80. Poplawski, A. & Binder, H. Feasibility of sample size calculation for RNA-seq studies. *Brief. Bioinform.* (2017). doi:10.1093/bib/bbw144
81. Nabel, E. G. & Braunwald, E. A Tale of Coronary Artery Disease and Myocardial Infarction. *N. Engl. J. Med.* **366**, 54–63 (2012).
82. Iqbal, J. *et al.* Bioresorbable scaffolds: rationale, current status, challenges, and future. *Eur. Heart J.* **35**, 765–776 (2014).
83. Ong, D. S. & Jang, I.-K. Causes, assessment, and treatment of stent thrombosis—intravascular imaging insights. *Nat. Rev. Cardiol.* **12**, 325–336 (2015).
84. Epstein, A. J., Polsky, D., Yang, F., Yang, L. & Groeneveld, P. W. Coronary revascularization trends in the United States, 2001-2008. *JAMA* **305**, 1769–1776 (2011).
85. Langer, R. & Tirrell, D. A. Designing materials for biology and medicine. *Nature* **428**, 487–492 (2004).
86. Rizik, D. G., Shah, M. G. & Burke, R. F. First US experience following FDA approval of the ABBOTT vascular bioresorbable vascular scaffold for the treatment of coronary artery disease. *Catheter. Cardiovasc. Interv. Off. J. Soc. Card. Angiogr. Interv.* **88**, 899–901 (2016).
87. Serruys, P. W. *et al.* A Polylactide Bioresorbable Scaffold Eluting Everolimus for Treatment of Coronary Stenosis: 5-Year Follow-Up. *J. Am. Coll. Cardiol.* **67**, 766–776 (2016).

88. Kang, J. *et al.* Bioresorbable Vascular Scaffolds - Are We Facing a Time of Crisis or One of Breakthrough? *Circ. J. Off. J. Jpn. Circ. Soc.* **81**, 1065–1074 (2017).
89. Mackel, M. J., Park, J. Y., Karageozian, H. L. & Karageozian, V. H. Integrin Receptor Antagonists and Their Methods of Use. (2013).
90. Yuet, K. P. & Tirrell, D. A. Chemical Tools for Temporally and Spatially Resolved Mass Spectrometry-Based Proteomics. *Ann. Biomed. Eng.* **42**, 299–311 (2014).
91. Robinson, R., Barathi, V. A., Chaurasia, S. S., Wong, T. Y. & Kern, T. S. Update on animal models of diabetic retinopathy: from molecular approaches to mice and higher mammals. *Dis. Model. Mech.* **5**, 444–456 (2012).
92. Ambati, J., Atkinson, J. P. & Gelfand, B. D. Immunology of age-related macular degeneration. *Nat. Rev. Immunol.* **13**, 438–451 (2013).

Chapter 2. *In vivo* mouse retinal transcriptome response to hypoxia and ALG-1001 treatment

2.1 Introduction

In the industrialized nations, retinal diseases are the main cause of clinical blindness[1]. In the United States alone, Age-related Macular Degeneration (AMD), Diabetic Retinopathy (DR), and Diabetic Macular Edema (DME) affect nearly 15 million Americans combined[2,3]. In these diseases, pathogenic hypoxia[4,5] and inflammation[4–6] of the retinal tissue promotes varying amounts of tissue-damaging neovascularization, edema and cell death (DR/DME in **Figure 1.4** and AMD in **Figure 1.5**). The current standard of care uses antibodies that target the Vascular Endothelial Growth Factor (VEGF), a potent pro-angiogenic cytokine[7]. For many patients, disrupting the disease-associated VEGF signaling slows disease progression and helps improve vision[2,8]. Despite their successes, current anti-VEGF treatments also have major limitations: 1) antibodies are expensive to manufacture with the high cost passed on to patients[9], 2) antibody-based treatment requires monthly intraocular injections that burden patients and healthcare providers[10], and 3) therapy is ineffective in more than 25-50% of the patients depending on the disease[11,12].

ALG-1001 or Risuteganib is an investigational oligopeptide drug that shows promise as an alternative therapy for a number of retinal diseases[13]. In a phase I trial of 15 late-stage DME patients, eight subjects experienced significant vision improvement, while the remaining seven did not show further vision deterioration[14]. Importantly, several of the patients that

responded to the drug are also refractory to the standard anti-VEGF treatment, suggesting ALG-1001 may be able to benefit patients who currently have no therapeutic option. More recently, several larger-scale, phase II trials in DR and wet AMD have been concluded with similar promising results[15–17]. In all of the human clinical trials conducted so far, ALG-1001 showed an excellent safety profile, with no observation of major drug-related side effects ([14,17–19] and personal communication). Interestingly, clinical data also suggest the visual benefit of ALG-1001 may last up to 90 days, which could help alleviate the need for monthly doctor visits[20]. Additionally, the oligopeptide may be a more cost-effective option as it can be readily produced synthetically and in more stable formulation than antibodies.

ALG-1001 has also been studied in animal models of retinal neovascularization that resemble DR and wet AMD (Campochiaro lab, Johns Hopkins University (JHU), unpublished). Oxygen-Induced Retinopathy (OIR) mouse is one of the models tested and is widely-used to study neovascularization that matches the development in human DR patients[21,22]. In this model, one week old pups are moved to a hyperoxia chamber, where high oxygen tension regresses existing retinal vasculature and inhibits retinal vessel development that normally occurs in the first few weeks postnatal (**Figure 2.1**)[23]. After 5 days of hyperoxia conditioning, mice are returned to room air, where insufficient oxygen supply due to under-developed vasculature leads to tissue hypoxia and rapid retinal neovascularization. This process lasts for approximately two weeks and eventually ends when balanced oxygen supply and demand are restored in the tissue. Thus, *unlike* conventional laboratory practice that places biological specimen in low oxygen environment to induce hypoxia, the OIR

model uses hyperoxia treatment to disrupt developing and existing vessels and induce a simulated hypoxia condition once the animal returns to room air.

While often used to study pathology of retinal neovascularization, the OIR model has proven to be a useful model for testing anti-angiogenic therapeutics[24,25]. In a study of ALG-1001 in the OIR model, a significant 54% reduction in retinal vessel growth was observed after ALG-1001 injection relative to vehicle alone (Campochiaro lab (JHU), personal communication). Furthermore, the therapeutic effect is dose-dependent, suggesting there is a specific modulation of retinal neovascularization by the drug.

While both clinical trials and *in vivo* animal studies generated promising results, ALG-1001's mechanism of action requires further investigation (**Chapter 1.3**). In this study, we applied RNA-sequencing (RNA-seq) to measure the drug-induced transcriptome changes in OIR retina to study the drug's mechanism of action. RNA-seq is well suited for detecting effects of drug-tissue interaction, as it is both high-throughput and unbiased[26]. By analyzing the entire expressed transcriptome in the mouse model, we found ALG-1001 specifically suppressed angiogenic and inflammatory genes and pathways elevated by hypoxia. In addition, when compared to other drugs, we found ALG-1001 shares gene regulatory profile with drugs that block disease-related cell proliferation pathways. Intriguingly, we only observed the therapeutic effect in OIR mice, but not in control mice, suggesting the ALG-1001's effect is dependent on the disease state of the retina.

To support these analyses, we performed an in-depth analysis of the OIR retinal transcriptome change compared to non-hypoxia stimulated control. On top of the expected elevation of angiogenesis process, hypoxia-modulated gene, pathway, and transcription factor analyses collectively suggested a profound activation of inflammatory processes as well as a moderate suppression of neural signaling and metabolic processes. Intriguingly, these biological mechanisms also play important roles in retinal disease pathophysiology[27], but are not well appreciated in the current literature on OIR model (predominantly thought as a model of retinal angiogenesis[23,28]). In agreement, our transcriptome analyses suggest OIR mouse may be suitable to model a broader range of retinal diseases and disease processes. Altogether, these findings provide novel insight into disease-relevant responses to hypoxia in the mammalian retina and modulation of the activated processes by ALG-1001.

2.2 Materials and Methods

Animal model and RNA-seq sample preparation

Animal tissues were prepared by the Campochiaro group (JHU) following a well-established OIR protocol[29–31]. In brief, two litters of mouse pups and their mothers were reared in room air (RA) until Postnatal day 6 (P6), then in an elevated oxygen environment (75% oxygen tension) for 5 days (P7-P11), and later returned to RA on day P12 (referred to as "OIR mice"). As a control, two litters of mouse pups and their mothers were raised at RA throughout the study (referred to as "RA mice"). On P12, when OIR mice were returned to RA, all mice received intravitreal injection of 1uL vehicle in the left eye and 1uL 1% ALG-1001 injection

(equivalent to around 1-3mM concentration after complete vitreous diffusion) in the right eye. On day P17, all mice were euthanized and their retinal tissues dissected and stored in RNA-later (ThermoFisher) at 4°C for 2 days. The tissues were then briefly washed with phosphate-buffered saline and disrupted using micro-sized pestle for tissue (Kimble Chase) and with QIAshredder (Qiagen). Next, total RNA were extracted using RNeasy Mini Kit (Qiagen) and contaminating DNA removed using TURBO DNA-free kit (ThermoFisher). RNA quality was measured using Bioanalyzer (Agilent Genomics); all samples were found to have least 8.80 RNA integrity number. RNA-seq libraries were then prepared using NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs Inc.) and sequenced on the HiSeq 2500 (Illumina) to generate on average 12.5 million of single-end, 100 base pair reads per sample.

RNA-seq analysis

Quality control

FASTQ files generated by RNA-seq were submitted to *FastQC* for quality control[32]. All samples showed good sequencing quality (**Figure S2.1**) and were submitted for read alignment. Alignment was performed using *Tophat2*[33] to the *mm10* genome and transcriptome references with *Bowtie 2* sensitivity level set to very sensitive (alignment results summarized in **Figure S2.2A**). After read alignment, gene counts were quantified using *HTSeq-count* using the intersection-strict model[34] (quantification results summarized

in **Figure S2.2B**). *FastQC, tophat2* and *HTSeq-count* reports were visualized using *multiQC* software[35].

Principle Component Analysis (PCA)

PCA was used to visualize the high-dimensional dataset. To eliminate null values, all counts in *HTSeq-count*-derived count table were increased by one. Transcript per Million (TPM) values were then calculated to correct for varying sequencing depth and gene length[36]. To reduce sensitivity to the uncertainty in TPM of low-expressing genes, PCA excluded genes for which the TPM averaged over all samples in the analysis was $\langle TPM \rangle < 5$. Five genes that have sex-specific expression profile were also excluded (*Xist, Uty, Kdm5d, Eif2s3y,* and *Ddx3y*) based on inter-sex |log2-fold-change|>3.0 based on *edgeR* measurement. Finally, natural log of TPM values were used as input for PCA. Based on the PCA results, we observed one obvious outlier in the dataset. Further analysis of this sample's profile suggest the tissue collected may be contaminated with lens cells (level of Cryaa gene that codes for lens crystalline protein[37] is 60 times higher in this sample compared to the average of other samples). Both the outlier and the fellow eye samples were removed from the rest of the study.

EdgeR differential expression comparison

To identify modulated genes, it is beneficial to exclude genes that are not expressed or expressed at very low level (their inclusion reduces statistical power). Here, we retained expressed genes, defined as those with Count per Million (CPM)>1 in at least five samples

(the smallest number of biological replicates from the same condition and litter). CPM is calculated to normalize for sequencing depth[36]. Typically, around 14,000 genes remain after filtering. Next, *edgeR* was used to perform the differential expression comparison by applying the GLM model to block out expression variability from other confounding factors[38]. When analyzing for differences between OIR + Vehicle and RA + Vehicle, variability associated with sex and litter was blocked out. When analyzing for differential expression between ALG-1001 and vehicle (performed in either OIR or RA), animal-to-animal variability was blocked out. After statistical test, genes with False Discovery Rate (FDR) ≤ 0.05 were considered as modulated and were further separated into positive and negative fold-change gene lists (summarized in **Figure S2.7**).

Goseq enrichment analysis of modulated gene list

Functional analyses of the hypoxia or ALG-1001-modulated genes were performed with the *goseq* package[39], which mitigates gene length bias inherent to RNA-seq. Enrichment was evaluated for both Gene Ontology (GO) terms (biological processes, cellular components, and molecular functions)[40] and Kyoto Encyclopedia of Genes and Genomes (KEGG) biological pathways[41]. GO terms associated with genes were obtained from *biomart*[42] while KEGG pathways were obtained from the KEGG REST server using the *keggrest* package[43]. Enrichment was performed separately for the positive and negative fold-change gene lists for each comparison. GO terms were considered enriched if FDR<0.05. The GO enrichment results were visualized using *REVIGO*, a visualization tool that aggregates closely-related GO terms, with similarity level set to small and GO term size determined using the *UniPort*

Mus musculus database[44]. Selected GO terms were labeled on the *REVIGO* plot based on our perception of their biological relevance. KEGG pathways were considered enriched if Benjamini-Hochberg adjusted p-value<0.05. In addition, "Human Diseases" KEGG pathways were removed to improve clarity.

GAGE pathway enrichment

In addition to *goseq* that analyzes genes with statistical significant change, *GAGE*[45] was used as an alternative approach to identify modulated KEGG pathways. *GAGE* uses the measured log2-fold-change values of KEGG pathway genes to test if they are significantly different from log2-fold-change values associated with the background. Unlike *goseq*, this approach is capable of picking up many small but consistent shift in measured transcript level changes that may not be considered statistically significant by *edgeR*. Same as *goseq*, a KEGG pathway is considered enriched if FDR<0.05.

ERSSA

ERSSA was used to check whether the number of biological replicates used in the RNA-seq experiment is sufficient to identify a majority of modulated genes[46]. Since *ERSSA* currently does not support blocking, it was only applied to the OIR + Vehicle vs. RA + Vehicle comparison. Analysis was performed with |log2-fold-change| cutoff of 0.5 and 50 subsamples at each replicate level.

ISMARA

ISMARA[47] was used to identify the transcription factors that may be responsible for the observed transcript level changes in response to OIR (OIR + Vehicle vs. RA + Vehicle) and in response to ALG-1001 (OIR + ALG-1001 vs. OIR + Vehicle). Samples involved in these two comparisons were submitted separately to *ISMARA* to obtain enrichment Z-scores. Analysis was focused on the top 50 most statistically significant transcription factor motifs as ranked by Z-score. Separately, all samples were also submitted together to obtain target gene activity profiles across all four conditions.

Within the top 50 enriched motif groups, we surveyed their function using Genecard[48] and Google searches and identified those with hypoxia and disease-relevant functions. Next, we identified KEGG pathways that are enriched in the top 50 TF motifs. In this analysis, we identified the KEGG pathways associated with the top TF motifs and compared them to the background, which consists of KEGG pathways associated with all TF analyzed by *ISMARA*. We tested for any KEGG Pathways that are over-represented in the top 50 motifs compared to the background using exact binominal test and then adjust the p-values using Benjamini & Hochberg approach[49] with statistical cut off set at FDR<0.05.

Cell type composition analysis

From previously published mouse retina single cell RNA-seq (scRNA-seq) study[50], we derived a list of genes that are highly expressed in specific retinal cell types. Briefly, among the previously identified 39 retinal cell clusters, we selected cluster #1, 2, 20, 24-26, 34-39 representing 12 main retinal cell types. For amacrine and bipolar cells, multiple clusters were

identified in the single cell study and for this analysis, we only considered the most populated cluster. For each cell type, we isolated the top 20 genes previously found to be highly expressed in the particular cluster. From this list, any duplicates among the cell types were removed to generate the final cell type-specific marker gene list (**Table 2.4**). This gene list was then used for heat map gene level visualization and fold change analysis to identify likely cell type composition change by hypoxia. Only genes with ⟨TPM⟩>0.10 were considered for analysis. Cell types with >80% of markers with positive fold change are considered expanded, >80% of markers with negative fold change are considered depleted and not changed if neither. To validate these results, the cell type composition analysis was repeated for mouse brain cell markers obtained from Zeisel *et al.*[51] Supplementary table 1, and mouse heart and immune cell markers from Skelly *et al.*[52] figure 1D.

Enrichr

To identify other drugs that have similar effect on the transcriptome as ALG-1001, we submitted the list of 198 suppressed genes after drug treatment in OIR mice to *Enrichr*[53] and compared them to the results of the LINCS L1000[54] and DrugMatrix datasets[55]. For this analysis, we looked for drugs that also suppressed the same genes as ALG-1001 with statistical significance (*Enrichr* adjusted p-value<0.05). For each analysis, we focused on the top 20 drugs with the strongest statistical evidence and identified their mechanism of action.

Enrichment analysis of Ishikawa et al. (2015) gene set

List of modulated genes in DR retinal tissue compared to non-diabetic retina were obtained from the original study[56]. GO biological process enrichment was performed with PANTHER[57] with adjusted p-value cutoff of 0.05.

Analysis of Ishikawa et al. (2010) data set

List of modulated genes in OIR retina compared to control retina were obtained from the original study[58]. Fold changes were calculated from the original microarray data using *GEO2R*[59] for plotting and Pearson's correlation calculations.

Analysis of Mirabelli et al. (2017) data set

With the Mirabelli *et al.* dataset[60], differential analysis was performed using *GEO2R* between anti-VEGF antibody and anti-IgG antibody treated samples. Fold changes were used for plotting and Pearson's correlation calculation.

Correlation measurement with DR and AMD studies

Hypoxia-modulated genes found in OIR + Vehicle vs. RA + Vehicle comparison were selected and converted to human gene symbol using *biomart*[42]. Fold changes for Ishikawa[56], Newman[61] and Whitmore[62] studies were calculated from the original microarray data using *GEO2R* for plotting and Pearson's correlation calculations.

Data availability

Raw RNA-seq Fastq files from this study will be available at Gene Expression Omnibus[63] once the study has been published. All enrichment tables will be made available at CaltechDATA[64].

Reverse transcription and quantitative polymerase chain reaction (qPCR)

From the statistically significant genes identified by *edgeR*, six genes were selected for separate validation with qPCR: Vegfa, Tgfbi, Col4a1 and Nrp1 (angiogenesis), Ccnb1 and Pdgfb (cell proliferation). SuperScript IV Reverse Transcriptase (ThermoFisher) and oligo(dT)20 primer (ThermoFisher) were used to convert polyadenylated mRNA to cDNA. qPCR was performed using iTaq Universal SYBR Green Supermix (Biorad) on LightCycler 480 (Roche) with 45 amplification cycles. Three animals from each of the two litters for each condition were randomly selected to constitute 6 biological replicates per condition. To enhance accuracy, four technical qPCR replicates of the same cDNA were generated and the median value selected for further analysis. Differences were calculated using the $\Delta\Delta C_t$ method versus *Gapdh*[65]. Primers used in this study were obtained from PrimerBank (**Table 2.1**)[66].

2.3 Results

In this study, RNA-seq was used to characterize the transcriptome in retinal tissue obtained from four conditions (RA and OIR mice treated with vehicle in the left eye and ALG-1001 in the right eye). First, we visualized the quantified profile of RNA-seq samples with PCA

to identify outliers (removed from subsequent analysis) and to observe sample separation among the four conditions. Next, we focused on functional analysis of the transcriptome-level response associated with retinal hypoxia and after ALG-1001 injection in both OIR and RA mice. Lastly, we applied analytical tools to identify transcription factors that likely influence the observed transcriptome changes.

PCA of RNA-seq data

PCA enables visualization of the RNA-seq dataset in a low-dimensional space to identify potential outliers and to give an initial impression on the strength of differences among the four conditions. First, we focused on the top three principal components ($PC1\text{-}3_{All}$) obtained from analysis of all 48 samples (12 per condition), which capture a combined 63% of the total variance in the dataset. $PC1_{All}$ alone accounts for 46% of total variance and clearly separates the OIR and RA samples (**Figure S2.3B**). Interestingly, one sample is a strong outlier in $PC2_{All}$ dimension (characterized in **Methods**, **Figure S2.3B**). To remove the outlier's effect on PCA, both the outlier and its fellow eye sample were removed and PCA repeated. The first three PCs from the revised PCA ($PCA_{No\text{-}outlier}$) again capture 62% of total variance (**Figure 2.2**). In addition, $PC1_{No\text{-}outlier}$ continues to strongly separate the OIR and RA samples, capturing 49% of the total variance in the data set (**Figure 2.2B**). In comparison, $PC2_{No\text{-}outlier}$ captures 7% of the total variance and appears to weakly separate the different litters, indicating the presence of litter effect in the dataset (**Figure S2.4A-B**). We also looked for PCA separation based on gender and found no observable pattern in $PC1\text{-}3_{No\text{-}outlier}$ (after removal of five gender-specific genes as described in **Methods**, **Figure S2.4C-D**).

In the analysis of the four groups, none of the top 3 PCs in $PCA_{No\text{-}outlier}$ separated vehicle and ALG-1001 conditions at room air, whereas there is a weak separation along $PC3_{No\text{-}outlier}$ among vehicle and ALG-1001 samples in OIR mice (**Figure 2.2C**). Based on this observation, PCA is repeated with only OIR samples to better visualize the quantified variation induced by ALG-1001 treatment (**Figure S2.5**). Some of the main observations from this analysis include: a) the strongest principle component only explains 17% of the variance among the samples; b) of the top three principal components, $PC2_{OIR}$ distinguishes vehicle from ALG-1001 (all mice show higher value on $PC2_{OIR}$ for vehicle relative to ALG-1001 treated fellow eye, **Figure 2.3A**). These results show ALG-1001 treatment induced a consistent perturbation of the transcriptome in the OIR mice. In contrast, no consistent shift was observed in $PC1\text{-}3_{RA}$ among the RA samples (PCA_{RA} in **Figure S2.6,** animal-specific analysis in **Figure 2.3B**).

Hypoxia-associated changes in the transcriptome

EdgeR was used to identify the genes that differ significantly in quantified level between hypoxia (OIR) and control (RA) retina, limited to vehicle samples to avoid confounding effect of ALG-1001. *EdgeR* analysis reveals 6208 genes modulated by hypoxia (satisfied FDR<0.05), representing more than 40% of the 14,402 expressed (⟨CPM⟩>1) genes (**Figure 2.4A**). The vast majority of the genes modulated by hypoxia also show very strong statistical significance ($FDR<<10^{-3}$, **Figure 2.4B**). Based on *ERSSA* analysis, a majority of hypoxia-modulated genes with |log2-fold-change|>0.50 have been identified (discovery becomes marginal past n=9, **Figure S2.8A**). In addition, *ERSSA* simulation shows consensus among

subsamples (**Figure S2.8B**), FPR and TPR (**Figure S2.8C**) metrics continue to improve as sample size increases, indicating enhanced accuracy as more samples are included.

For functional analysis of the genes that are modulated by hypoxia, they are separated into two lists (3141 "up" and 3067 "down") to identify biological processes (GO) and pathways (KEGG) that are over-represented. Beyond "response to hypoxia", GO biological process analysis of elevated genes disproportionally identified those that are associated with inflammation ("immune system process", 326 out of 738 expressed "immune response" genes are elevated) and angiogenesis ("blood vessel development", 181 out of 374 expressed "angiogenesis" genes are elevated) (**Figure 2.5A**), which are qualitatively in agreement with prior literature (**Appendix 2.8.1**). A number of functions identified are relevant to both angiogenesis[67] and inflammation[68], including "response to stimulus", "cell proliferation", and "cytokine production". Additionally, a significant portion of hypoxia-elevated genes are strongly activated, including 586/3141 genes that are increased by more than one-fold (**Figure S2.7**). Intriguingly, processes related to glycolysis that were enriched in prior study with shorter hypoxia exposure (1d, P13) are not enriched in this analysis (5d, P17), suggesting resolved activation of glycolysis by P17 (**Appendix 2.8.1**).

In contrast, GO biological process analysis of hypoxia-suppressed genes disproportionally identified processes related to neural system, such as "ion transport", "chemical synaptic transmission", and "neuron projection development" (**Figure 2.5B**). Compared to hypoxia-elevated genes, the effect on the suppressed genes is more modest as only 78 genes are reduced by more than one-fold (**Figure S2.7**). Perhaps due to these small changes and the

longer hypoxia exposure, transcriptome studies of shorter hypoxia exposure in OIR mice did not find suppression of neural genes (**Appendix 2.8.1**).

Beyond the hypoxia-response "HIF-1a signaling pathway"[69], KEGG pathway analysis using *goseq* and the genes elevated in OIR identified cell proliferation and migration pathways that are involved in angiogenesis and inflammation: "PI3k-Akt signaling pathway"[70,71], "focal adhesion"[72,73], "regulation of actin cytoskeleton"[71,74], and "cell cycle"[75,76] (**Figure 2.6**, *goseq*). Additional KEGG pathways mainly associated with inflammation[77] are also enriched, including "cytokine-cytokine receptor interaction", "NF-Kappa B signaling pathway", "JAK-STAT signaling pathway", "complement and coagulation cascades", and "leukocyte transendothelial migration". In contrast, *goseq* applied to hypoxia-suppressed genes in OIR identified KEGG pathways that predominantly belong to two groups: metabolism and neural systems (**Figure 2.7**, *goseq*). Metabolism pathways suppressed by hypoxia include "oxidative phosphorylation" and "citrate cycle (TCA cycle)", which are responsible for aerobic respiration[78]. In relation to neural system, a large number of related KEGG pathways are suppressed by hypoxia, including those that function in signaling, synapse, and secretory processes[79].

In addition to *goseq* (a statistical analysis of annotated gene set over-representation in a list of modulated genes), *GAGE* (a statistical analysis of rank order of annotated gene set among expressed genes ordered by fold change) was used as an alternative method of identifying KEGG pathways using observed transcriptome changes. When applied to analyze fold changes affected by hypoxia, *GAGE* identified fewer KEGG pathways than *goseq* and only

one ("TNF signaling") that is not discovered by *goseq* (**Figure 2.6 and 2.7**, *GAGE*). Altogether, *goseq* and *GAGE* analysis provide statistical evidence that angiogenesis and inflammation-related pathway genes are elevated, while metabolism and neural system-related pathway genes are suppressed.

Transcriptome modulation by ALG-1001

As evident in the PCA analysis, the effects of ALG-1001 treatment are measurable in OIR mice (albeit much weaker than the effects of hypoxia, a weak separation in $PC3_{No\text{-}outlier}$ in **Figure 2.2C,** consistent shift in $PC2_{OIR}$ in **Figure 2.3A**), but not in RA mice (No separation in PC1-$3_{No\text{-}outlier}$ in **Figure 2.2B-C,** no shift in PC1-3_{RA} in **Figure 2.3B**). In contrast to the effects of hypoxia, which involved thousands of modulated genes (both "up" and "down", **Figure 2.4A**), ALG-1001 treatment in RA (**Figure 2.8A**) and OIR (**Figure 2.8B**) mice affected only a few hundred genes, and essentially all with negative fold change (95% and 94% of all ALG-1001-modulated genes in RA and OIR mice, respectively).

Next, *goseq*-based functional analysis was used to study the suppressed genes (the number of elevated genes was too small for meaningful functional enrichment). When applied to the 70 genes suppressed after ALG-1001 treatment in RA mice, *goseq* found a limited number of GO biological processes with weak statistical significance (**Figure 2.9A**). Curiously, vision-related processes such as "retinol metabolism", "visual perception", and "sensory perception of light stimulus" are enriched. In particular, five genes involved in retinol metabolism (Lrat, Rbp1, Rdh5, Rpe65, and Ttr) are suppressed after ALG-1001 treatment in

RA mice. Since retinol plays an important role in the detection of light by photoreceptors[80], it is unclear whether moderate suppression of these genes may alter normal vision function.

Analysis of 198 suppressed genes after ALG-1001 treatment in OIR mice identified many GO biological processes discovered with hypoxia-elevated genes (89/115 of enriched processes), including many related to angiogenesis[81,82]: "extracellular matrix organization", "integrin-mediated signaling pathway", "circulatory system development", and "tube development" (**Figure 2.9B**). While not directly shown on the *REVIGO* plot, "angiogenesis" is also enriched, including 27 of 198 ALG-1001-suppressed genes (many of which are also elevated by hypoxia, including Vegfa). Additionally, cellular components (**Figure S2.9A**) and molecular functions (**Figure S2.9B**) analyses showed strong evidence that genes encoding extracellular matrix-located and structural-related proteins are suppressed after ALG-1001 treatment in OIR mice. In particular, many of these proteins form the extracellular matrix or are proteins that bind to the matrix, both of which are important for cell adhesion and migration during angiogenesis[83].

Goseq-based KEGG pathway enrichment supported the results from GO biological process analysis. In RA mice, "retinol metabolism" and two other KEGG pathways are enriched with ALG-1001-suppressed genes (**Figure 2.10B**, *goseq*, RA mice). However, evidence for their modulation are weak as indicated by the large adjusted p-values. On the other hand, four KEGG pathways are enriched with ALG-1001-suppressed genes in OIR mice, all of which are also enriched with hypoxia-elevated genes (highlighted, **Figure 2.10B**, *goseq*, OIR mice). Of the four, "PI3K-Akt signaling pathway", "ECM-receptor interaction", and "Focal

adhesion" are associated with various aspects of the angiogenesis and inflammation processes[70–73]. Within the last pathway – "Protein digestion and absorption", we find ALG-1001 suppressed a cluster of collagen genes that are involved in both protein metabolism and cell adhesion. Since the retinal tissue is not involved in gastrointestinal-based protein metabolism, enrichment of this pathway appears to be due to the diverse roles that collagens are involved in.

GAGE analysis of KEGG pathway enrichment corroborated the *goseq* findings, improving the confidence of pathway identification and revealed a few additional modulated pathways related to ALG-1001 treatment in OIR mice (**Figure 2.10**¸*GAGE*). The four KEGG pathways that are identified by *goseq* with ALG-1001-suppressed genes in OIR mice are indeed the most confidently identified pathways found by *GAGE* (adjusted p-value<10^{-5}, **Figure 2.10B**, *GAGE*, OIR mice). In addition, *GAGE* detected four additional KEGG pathways related to angiogenesis and inflammation and all eight pathways are enriched with hypoxia-elevated genes (highlighted in **Figure 2.6**, *GAGE*). Significantly, the overlap with hypoxia-modulated pathways is only observed with ALG-1001 treatment in OIR mice, but not in RA mice (compare **Figure 2.10B**, *GAGE*, OIR mice and **Figure 2.10B**, *GAGE*, RA mice).

One of the main advantages of *GAGE* is it can detect small, but consistent shifts in the transcriptome among genes of a specific annotated gene set. In some cases, changes in quantified transcript level may be too small or noisy to be identified as statistically significant by software such as *edgeR*, but these small shifts can collectively have a meaningful biological effect. In this case, ALG-1001 only elevated a few genes with statistical

significance, which precludes *goseq* identification of related KEGG pathways. However, *GAGE* found two translation-related KEGG pathways that are elevated in response to ALG-1001 in OIR mice ("Spliceosome" in **Figure 2.10A**, OIR mice) and in RA mice ("Ribosome" in **Figure 2.10A**, RA mice). While enrichment of the two pathways are not supported by *goseq*, these findings suggest ALG-1001 treatment elevated genes in protein synthesis pathways[84], which are known to be disrupted in retinal degeneration[85].

Transcription factors enrichment

Utilizing the transcriptome data, we used *ISMARA* to search for the transcription factors (TFs) that are likely involved in influencing the quantified changes in response to hypoxia (OIR + Vehicle vs. RA + Vehicle) and to ALG-1001 treatment in OIR mice (OIR + Vehicle vs. RA + Vehicle). The result is in form of TFs ranked by an enrichment Z-score that measures the confidence of their involvement (hypoxia response in **Table 2.2** and ALG-1001 response in **Table 2.3**). *ISMARA* shows many TFs that likely elevated target genes after hypoxia, including known regulators of cell proliferation (Jun, E2F, ETS, and Tead family members), angiogenesis (Fos), and inflammation (Stat2, Spi1, Irf family members, Nfatc3, and NF-κB family members) (labeled in **Figure 2.11A**, activity profile in **Figure 2.12**)[48,86]. Similarly, TFs with suppressed target genes after hypoxia include those relevant to cell proliferation (Hoxb7, Pml, Etv1, and Zbtb14), inflammation (Etv3), neuronal development (Rest), metabolism (Ppara), and hypoxia-response (Arnt) (labeled in **Figure 2.11B,** activity profile in **Figure 2.13**)[48].

Consistent with earlier analyses, enrichment for TFs corresponding to the transcriptome change after ALG-1001 treatment in OIR mice is comparatively less significant. The difficulties posed by the small number of the modulated genes after drug treatment are particularly acute for *ISMARA*, which currently does not support removal of inter-animal variation to improve detection. As a result, none of the top 50 TFs have Z-score>1.0, which is generally used as the significance cutoff (**Table 2.3**). Nevertheless, biologically relevant TFs are observed in the top 50 TFs (labeled in **Figure 2.14**), with some evidence that indicate opposing effect of hypoxia and ALG-1001 (ALG-1001 reversed hypoxia-modulated activity level in 8/12 of biologically relevant TF motifs in **Figure 2.15**). Consistent with *edgeR*, majority of these TFs suppressed target gene activity, including several cell fate and cell proliferation-regulating regulator (Erg, Fosl2, Fos, Id4, Sox14, Jun, and Trp53), a hypoxia-response regulator (Epas1), and two Bcl TFs (Bcl3 and Bcl6) that regulate inflammatory and cell-survival processes (**Figure 2.15B**)[48]. In particular, Erg, Fos, and Jun are also previously enriched with elevated target genes after hypoxia. At the same time, the top 50 TF groups also include three biologically relevant TFs with elevated target genes after ALG-1001 treatment, including Etv1 and Etv6 (cell growth, angiogenesis, migration, and proliferation), and Foxo3 (apoptosis and protection against oxidative stress) (**Figure 2.15A**).

qPCR

qPCR was used to validate RNA-seq measured gene level profile for a limited number of genes. For this purpose, we selected six genes that are elevated by hypoxia and suppressed by ALG-1001, and are involved in angiogenesis (Vegfa, Tgfbi, Col4a1, and Nrp1) or cell

proliferation (Ccnb1 and Pdgfb) processes[48]. Consistent with RNA-seq (**Figure 2.16A**), qPCR measured changes showed the same pattern across all six genes (**Figure 2.16B**), demonstrating high reproducibility between the two detection methods.

2.4 Discussion

Hypoxia and the associated biological response play central roles in developmental processes, injury recovery and chronic diseases (as detailed in reviews[87–89]). In the eye, a pathological hypoxia condition is associated with diseases such as DR[90], wet AMD[91], retinopathy of prematurity[90], and intraocular tumor[92]. In DR, hyperglycemia leads to oxidative stress and inflammation, which cause dysfunction of retinal vasculature and consequently, hypoxia (**Figure 1.4**)[93,94]. In AMD, the inability of the retinal pigment epithelium (RPE) to keep pace with the need for phagocytic maintenance of the posterior retina leads to development of waste deposits and inflammation, which cause failure of the blood-retina barrier and consequently, hypoxia (**Figure 1.5**)[91,95,96]. Subsequently, inadequate oxygen supply promotes abnormal retinal angiogenesis and inflammation that contribute to tissue damage and vision loss.

As an unbiased source of information regarding the hypoxia response in OIR mouse retina, the value of RNA-seq comes through in four ways: a) it reveals new insight into the possible relationship between this animal model to human retinal diseases, b) points to possible therapeutic targets beyond the current clinically-approved therapies, c) shows evidence of change in subpopulations of retinal cells during the hypoxia response, and d) demonstrates

how the effects of a proposed therapeutic relate to the modulation of the transcriptome associated with hypoxia.

The RNA-seq results give evidence that the OIR model may have broader relevance than the conventional view that it only simulates angiogenic conditions in DR patients: the transcriptome suggests that the elevation of the inflammatory response in the OIR mouse retina may make it a valuable model of retinal inflammation in both DR and AMD. In relation to angiogenic conditions, quantitative information from RNA-seq points to targets that are potentially more potent than anti-VEGF therapies that are currently the only clinically approved drugs. Applying RNA-seq to an intact tissue (here the retina) reflects two types of changes in the transcriptome: changes in tissue composition as result of proliferation, death and migration of certain types of cells each with a characteristic expression profile and changes in expression of certain genes by cells in the tissue. In relation to an investigational drug, application of RNA-seq in a four-arm study reveals that a particular compound has very little effect on the retina of unstressed mice and has effects that reverse many of the responses to hypoxia when administered to OIR mice.

Retinal hypoxia activates angiogenesis and inflammation

To better understand the biological processes active in the retina when insufficient oxygen is available, we use the OIR mouse model that exhibits aberrant retinal angiogenesis after suppression of retinal vasculature development by raising pups in a hyperoxia chamber (**Figure 2.1**)[21,23]. In the retina research community, the OIR model is widely used to study

retinal angiogenesis[23], but considerably less attention has been placed on inflammation and other processes activated by hypoxia[27]. In this study, at height of morphological angiogenesis (5 days after inception of simulated hypoxia by returning pups to a normoxic atmosphere)[97], we found significant elevation of not only angiogenesis, but also inflammatory genes in the OIR retina compared to control retina: GO analysis identifies "immune system process" and "blood vessel development" (*goseq* and *REVIGO* in **Figure 2.5A**), while enrichment analysis identifies KEGG pathways related to angiogenesis (*e.g.* "PI3K-Akt signaling pathway"[70,71] and "focal adhesion"[72]) and inflammation (*e.g.* "NF-Kappa B signaling pathway"[77] and "leukocyte transendothelial migration"[98]) (*goseq* in **Figure 2.6**, and *GAGE* in **Figure 2.6**). More broadly, KEGG pathways enriched with genes elevated in OIR retina are involved in diverse cellular and system functions including adhesion, migration, proliferation, cell death, and inflammation. More importantly, many of the enriched pathways are implicated in DR and AMD human pathophysiology (*e.g.* "PI3K-AKT"[99,100], "MAPK signaling" [99,100], "NF-kappa B"[101] and "JAK-STAT"[102] pathways).

Corroborating GO biological process and KEGG pathway analysis of hypoxia-elevated genes, TF enrichment also strongly points to transcriptome regulation of both angiogenesis and inflammation in the OIR retina (labeled TFs in **Figure 2.11**). In particular, TFs associated with both processes are over-represented with statistical significance in the list of top 50 enriched TF motifs (**Appendix 2.8.11**). A particular strength of TF analysis is that it is capable of implicating TFs that are *uncharacterized* altogether, much less having known roles in the hypoxia response. TFs that are strongly enriched and not yet annotated (*e.g.*

Nhlh1 and Hsfy2) may represent candidates for future research in retinal hypoxia biology (**Appendix 2.8.11**).

Retinal hypoxia disrupts neural retina

Enrichment analysis of genes that are suppressed in OIR retina pointed to cellular metabolism and neural system processes (GO biological processes, **Figure 2.5B**). In relation to metabolism, enriched KEGG pathways include aerobic "oxidative phosphorylation" and "citrate cycle (TCA cycle)" (**Figure 2.7**), consistent with prior literature reporting a transition away from aerobic respiration to glycolysis in OIR mouse[103] and in human AMD[104] and DR[105,106] patients. In relation to neural processes, GO biological process enrichment strongly identified neural system processes, including "synapse assembly", "neurotransmitter transport", and "neuron projection development" (**Figure 2.5B**). Similarly, genes in multiple neurotransmitter signaling and neuronal synapse KEGG pathways are suppressed (*goseq* and *GAGE* in **Figure 2.7**), likely due to disruption of normal neural development by hypoxia-activated processes in OIR retina (in mouse pups, significant retinal neuron differentiation and synapse formation take place during the first 3 weeks postnatal[107–109]). Among the top 50 TFs identified by *ISMARA*, only 1/3 were depressed in OIR; nevertheless, this small group includes the only neural and metabolism-related TF identified: Ppara (regulator of metabolism) and Rest (regulator of neuronal genes) (**Figure 2.11B**). Although the TF evidence is weak, it is compatible with the GO biological process and KEGG pathway evidence that metabolism and neural system processes are suppressed in the OIR retina.

OIR mouse as a model of human retinal diseases

The human retina is supplied by two types of vasculature: blood vessels that reside within the retina and the choroid capillaries that lie outside of RPE and Bruch's membrane[28] (**Figure 2.23**). Physiologically, neovascularization arises from different locations in wet AMD and DR. In wet AMD, vessels grow from the choroidal region. In DR, angiogenesis originates from the retinal vessels. Due to these differences, several animal models have been created to mimic each condition[23]. The mouse OIR model develops new vessels from the retinal blood vessels, as is the case in DR and unlike wet AMD. Therefore, most of the literature on the OIR mouse model asserts that it is a useful model for DR (bearing in mind the significant differences between the mouse model and the human retinal diseases, **Appendix 2.8.12**), and the conventional wisdom that the OIR mouse is not suitable to model AMD. The present RNA-seq data give us an opportunity to compare the changes in the retinal transcriptome of OIR mouse to those observed in human donor retina of particular eye diseases relative to individuals with no record of retinal disease.

We begin by comparing the transcriptome changes in OIR to those reported in human donors diagnosed with DR in two prior transcriptome studies[56,110]. A study of human retinal vascular endothelial cells derived from DR patients vs. non-diabetic individuals[110] reported enrichment of 57 GO biological processes among regulated genes – 42 (74%, p-value=0) of which were found in the present study of OIR retina, include "angiogenesis", "inflammatory response", "regulation of cell death", and other related processes. Another study of human DR patients (DR vs. control retina)[56], identified 87 elevated genes that are used to enrich 39

GO biological processes – 30 (77%, p-value=0) of which overlap our study, with most that are related to angiogenesis and inflammation. Intriguingly, we also found some evidence that suggest suppression of neural genes in DR retina, consistent with our observation in OIR mouse (**Appendix 2.8.3**).

Next, we performed a quantitative analysis that compares transcriptome profile of hypoxia-modulated genes in OIR retina to those found in DR and AMD patients. Based on prior literature, we expected to see a stronger correlation between mouse OIR and human DR, than for mouse OIR and human AMD. Indeed, hypoxia-modulated gene changes correlated more strongly with DR (Pearson's r=0.3346, **Figure 2.24A**) than to AMD (r=0.0459-0.2606, depending on the stage of AMD, **Figure 2.24C**). We are surprised to see indications that the mouse OIR model captures the immune system process differences between diseased and control human retinas as well as angiogenesis genes: when only "immune system process" genes are considered, the correlation between OIR mouse and human DR is significant (r=0.3948) and nearly equal to the correlation of "angiogenesis" genes (r=0.4073). Indeed, the OIR mouse model correlates with immune system processes in AMD, particularly for disease states that are not associated with angiogenesis: r=0.261, 0.231 and 0.282 for immune system process genes in sub-clinical, dry, and GA AMD, respectively. For these three forms of AMD, hypoxia-modulated genes associated with cell death and neuron system processes also showed significant correlation with OIR mouse (**Appendix 2.8.4**).

Over the past 20 years, the mouse OIR model has been widely used to study retinal angiogenesis, but the field has only recently started to appreciate the concurrent

inflammatory response to hypoxia[27]. In the broader research community, the co-occurrence of hypoxia-induced angiogenesis and activation of inflammation have been reported in a variety of human diseases[111,112]. Mechanistically, oxygen deprivation activates angiogenesis to increase oxygen supply[82] and creates a local microenvironment that promotes inflammation[113] and cell death[114,115]. Additionally, there is positive cross-talk between the activated angiogenesis and inflammation: early angiogenesis forms leaky vessels that enhance leukocyte adhesion and migration into the tissue to promote inflammation[116], while inflammation increases oxygen demand in the tissue and activates angiogenesis[117]. As a result, dysregulation of these processes are thought to reinforce each other to create a persistent cytotoxic environment in chronic human diseases such as DR and AMD[111]. In relation, our comparative analysis suggest OIR mice may be relevant to a broader range of human retinal diseases (both DR and certain stages of AMD) and disease processes (angiogenesis, inflammation, neural disruption, and cell death).

Angiogenic targets beyond VEGF

Despite substantial clinical investigation in the past several decades, the only class of drug currently approved for DR[118] and wet AMD[119] consists of antibodies that target VEGF. VEGF is a potent pro-angiogenic factor that has been extensively studied in the past for its prominent role in activating angiogenesis[120], including in the OIR mouse model[21]. In an examination of elevated angiogenesis genes in OIR retina at P17, we found one third of them (61 out of 181 total) are quantified to be elevated more than Vegfa (encodes pro-angiogenic form of VEGF, 2.27 fold increase, FDR=4.30×10^{-30}). Functionally, these 61 genes are

closely related (**Figure S2.10**), with many involved in cell adhesion/migration (*e.g.* integrins, collagens, fibronectin, and cadherin) and cell proliferation (*e.g.* growth factors, tyrosine protein kinases and their associated proteins), both processes that are essential during angiogenesis (**Table 2.8**). Significantly, only 13 of these genes are downstream of VEGF signaling (**Table 2.9**). Since a majority (approximately 80%) of these highly-elevated angiogenesis genes do *not* appear to be part of VEGF signaling, they suggest other concurrent processes are important in driving angiogenesis in the hypoxic retina. Interestingly, Placental Growth Factor (Pgf)[121] and Fibroblast Growth Factor 2 (Fgf2)[122] are both potent pro-angiogenic factors found to be elevated more than Vegfa, but have not been extensively tested in clinical trials of retinal diseases (none that specifically targets Pgf [123] or Fgf2[124]). Therapeutics targeting these growth factors and other highly elevated angiogenic proteins may lead to novel treatment of retinal diseases.

Modulation of retinal cell composition by hypoxia

Unlike traditional *in vitro* cell culture, animal tissue is comprised of multiple cell types whose populations are not static. In the retina, hypoxia response initiates subpopulation change among the retinal cells, which can influence the transcriptome measurement and contribute to differential expression identification. Specifically, hypoxia activates angiogenesis and inflammation that elevate blood vessel[125,126] and immune[127–130] cell populations in the retina through proliferation of resident cells and migration of systemic cells into the tissue. Additionally, hypoxic retinal cells increase production of reactive oxygen species and release excess glutamate that induce neuronal cell death[131]. Prior studies that used flow cytometry[129]

or fluorescent staining[125,126,128] methods did observe changes in cell population in the OIR model, but these methods are low-throughput with the studies focused on a limited number of cell types at a time. On the other hand, while RNA-seq data contain transcriptome signal combined from all retinal cell types, it also allow us to analyze all of the retinal cell types at once.

To investigate the representation of 12 common retinal cell types, we used lists of mutually orthogonal gene sets that we derived from a recent scRNA-seq study of the mouse retina[50].

Using these cell-type specific gene sets (as described in **Methods**), we found evidence that 3 cell types may be expanded (**Figure 2.19A**), 3 cell types depleted (**Figure 2.19B**), and 6 cell types with no strong evidence for change in cell population after hypoxia (**Figure 2.20**). In particular, we found vascular endothelial cells, pericytes and microglia numbers are likely expanded in OIR retina (supported by additional evidence using orthogonal markers from a brain[51] study and a heart[52] study, **Appendix 2.8.2**). Significantly, endothelial cells and pericytes are the main cells that form capillary vessels, while microglia are the retina-resident immune cells. Evidence for their expansion suggest the observed elevation of angiogenesis and immune genes in the OIR retina is not due to only gene expression change, but also tissue composition variation.

We also identified three neuronal cell types that may have depleted population in OIR retina: ganglion, amacrine and bipolar cells, which represent 3 out of 4 major neuronal cell types in the retina[132] (supported by additional evidence using orthogonal markers from a brain study,

Appendix 2.8.2). Consistent with these findings, prior study of OIR retina found evidence of neuron cell death (in layers of retina containing amacrine, bipolar, horizontal and ganglion cells) at P16-17[133,134]. However, imaging studies also found photoreceptors are susceptible to undergo cell death, which was not strongly detected in our transcriptome dataset (rods 11/20 markers, cones 5/13 markers with negative fold-change in OIR mice). This discrepancy between imaging and transcriptome studies may be due to the significantly larger number of photoreceptors in the retina[135], where death of a limited number of photoreceptors is not detectable by whole retina RNA-seq.

Advantage of drug testing in both stressed and control animals

Despite over 2000 clinical trials to date[136], anti-VEGF therapy is the only class of drug approved for DR and wet AMD[137], and alarmingly do not work for double digit percentage of patients (**Chapter 1.1**). Despite tremendous amount of clinical efforts, most experimental drugs fail in these diseases due to a lack of efficacy or presence of strong adverse events[138]. Thus, any new therapy needs to show both strong therapeutic benefit in patients while retaining a high safety profile. ALG-1001 is an experimental drug that matches this profile with demonstration of: a) vision improvement across multiple clinical trials in DME and AMD patient cohorts[15–17] and b) no strong drug-related adverse events after more than several hundred human injections (Allegro Ophthalmics, LLC., personal communication). In this study, RNA-seq is applied to investigate ALG-1001's safety profile in control mice (RA) and therapeutic effect in disease model mice (OIR).

Inclusion of both the unstressed group and the disease model enables us to discover a striking difference between effects of an active agent in the two treatment conditions. Specifically, we discovered ALG-1001's pronounced anti-angiogenic and anti-inflammatory effects in the OIR group are absent in the RA group (compare GO biological processes enriched in RA group in **Figure 2.9A** and OIR group in **Figure 2.9B**). Indeed, the ALG-1001 treated eyes in the control animals are almost indistinguishable from their sham treated fellows (number of drug modulated genes in **Figure 2.8A,** biological processes enriched in **Figure 2.9A** and lack of separation in PCA_{RA} in **Figure 2.3B**). Altogether, this two-pronged experimental design enables quantitative identification of the (potentially therapeutic) effects of an active agent in a diseased tissue and the (desirable) absence of effects in unstressed tissue. Our findings show potential benefits of using this type of experimental design, which is uncommon in the research community that uses transcriptome methods to study drugs and diseases (small-scale survey with results summarized in **Appendix 2.8.9**).

ALG-1001 modulates hypoxia-induced transcriptome changes

After ALG-1001 treatment, transcriptome data reveal drug exposure in OIR mice suppressed genes that are part of hypoxia-activated angiogenesis and inflammatory-related biological processes (GO biological processes, **Figure 2.9B**) and pathways (*goseq* in **Figure 2.10B,** *GAGE* in **Figure 2.10B**), along with corroborating regulatory evidence (**Appendix 2.8.7**) and analysis that show a general inverse relationship between effects of hypoxia and ALG-1001 (**Appendix 2.8.13**). Expanding on these observations, a protein-protein interaction graph of the 198 drug-suppressed genes revealed two noticeable clusters of highly-

interacting proteins (**Figure 2.25**). One cluster contains growth factors (Vegfa, Fgf1 and Pdgfb), growth factor receptors (Met, Pgr and Fgfr2), cell cycle regulators (Ccnb1, Ccnd1, Cdk14 and Cdkn1a) and angiogenesis-related proteins (Nrp1, Ets1 and Esm1). The second cluster contains many highly related structural proteins including 8 collagen subunits, 3 laminin subunits, and others (Bgn, Itga9, Timp3, and Zyx). In particular, collagen and laminin are important for cell adhesion and migration and promote angiogenesis by interacting with endothelial integrin receptors[83]. Focusing on integrins, we found only one integrin subunit (Itga9) is suppressed with statistical significance, but 16/19 of expressed integrin subunits have negative fold changes. These changes are too small for statistical significance, but when combined with suppression of their extracellular matrix (ECM) binding partners, the overall effect may explain drug's anti-angiogenic mechanism. Indeed, *in vitro* studies of endothelial cells treated with ALG-1001 revealed suppression of cell migration and adhesion genes (**Chapter 3**), which result in measurable reduction of physical adhesion and migration of endothelial cells on ECM-coated surfaces (Campochiaro lab, JHU, described in **Appendix 2.8.5**).

Moderation of hypoxia's transcriptome effect correlates with observation of mitigated OIR vessel growth after ALG-1001 treatment, in a manner that is comparable to current standard of care – anti-VEGF antibodies. In this study, we did not test anti-VEGF antibodies (due to the high cost of RNA-seq and the paired eye experimental design), but we set out to explore the existing literature on transcriptome studies of anti-VEGF treatment for comparative analysis with ALG-1001. Despite an extensive search, we only found one microarray dataset

that tested anti-VEGF therapy in a stressed rat cornea model and, due to the modest effect of anti-VEGF therapy in this model (no regulated genes found), the comparative assessment is inconclusive (**Appendix 2.8.6**). Next, we broadened our analysis to large-scale transcriptome studies of drugs, where *enrichr* was used to search for drugs that primarily suppressed the same genes as ALG-1001.

Despite wide adoption of high-throughput transcriptome methods, only a handful of studies had explored drug-induced transcriptome changes on a large number of drugs. Specifically, we found two datasets in which we could search for drugs that show significant similarity to ALG-1001: 1) the LINCS L1000 data set provides ~20,000 compounds/drugs studied in human cancer cell lines across around ~30,000 total experimental settings (varying dosage, treatment time, and cell lines)[54] and 2) DrugMatrix data set contains 600 compounds tested in rat across around 4,000 total experimental settings (varying dosage, treatment time, and analyzed tissue)[55]. First, with the L1000 dataset, *enrichr* found ALG-1001 transcriptome profile matches 390 compounds with statistical significance, albeit with important caveats (**Appendix 2.8.8**). Within the set of 390 drugs, we are surprised that one of ALG-1001's attributes is very rare: genes modulated by ALG-1001 overwhelmingly (93.8%) have negative fold change, but only 4 of the 390 drugs showing similarities to ALG-1001 have >80% of their DE genes in the negative fold-change category (none of 390 have >90% of DE gene as suppressed). Similarly, when this analysis is applied to the DrugMatrix dataset, we found ALG-1001 expression profile matches 89 drugs with statistical significance. Of the 89 drugs with similarities to ALG-1001, none have >70% of their modulated genes as

suppressed (only one drug showed >60% of their modulated genes have negative fold change). Collectively, these observations suggest ALG-1001 is relatively unique in terms of its ability to mainly suppress genes.

Next, we set out to identify any known mechanism of action among the top 20 enriched L1000 (**Table 2.11**) and DrugMatrix (**Table 2.12**) compounds. Through a manual search, we found known mechanism of action for a majority of the top 20 L1000 drugs, while the opposite is true with DrugMatrix compounds (**Appendix 2.8.10** describes annotation of DrugMatrix top 20 enriched compounds). Here, we will focus on the enriched L1000 compounds as they can help expand our understanding of ALG-1001's mechanism of action. Among these 20 compounds, we are intrigued to see a majority of them (18/20) are either cancer drugs or cell cycle inhibitors. When specifically focused on the cancer drugs, we notice more than half of them (8/15, bolded in **Table 2.11**) are inhibitors of tyrosine kinases, which are a large family of proteins that facilitate signal transduction in the cell and play key roles in pathological cell proliferation, differentiation, migration, metabolism and cell death[139,140]. In relation to these observations, specific enrichment of tyrosine kinase inhibitors suggest ALG-1001 may disrupt targets in related processes, potentially explaining its positive effect in retinal diseases.

2.5 Conclusions

Neovascularizing diseases of the retina are the main cause of clinical blindness in the developed nations[1]. In these diseases, hypoxia is one of the main drivers of

neovascularization and inflammation that subsequently disrupt cell function and cause cell death[91,93,95]. Current treatments for these conditions are expensive[9], burdensome[10] and alarmingly do not work in a double digit percentage of patients[11,12]. ALG-1001 is an investigational peptide drug that is promising in treatment of blinding retinal diseases[14,19], but its mechanism of action requires further exploration.

In this study, RNA-seq shows simulated hypoxia in mouse OIR retina modulated thousands of genes: strongly elevating those related to angiogenesis and inflammation, while weakly suppressing others related to metabolism and neural system processes. These quantified changes can be explained by a combination of gene expression change in retinal cells and hypoxia-induced tissue composition variation, as revealed in a robust cell marker analysis. Additionally, quantitative comparative analyses to prior human datasets suggest the OIR model may be useful to model a wider range of human retinal diseases and disease processes. More broadly, our comparatively large transcriptome dataset can serve as a resource to the retinal research community as a foundation to further unravel the extremely complex biological processes that drive retinal diseases.

In terms of therapeutic treatment of these diseases, RNA-seq reveal ALG-1001 co-treatment reversed of some of the transcriptome changes activated by hypoxia in OIR retina, specifically, the stimulated angiogenesis and inflammation processes. These findings support unpublished studies that showed ALG-1001 is effective in mitigating hypoxia-induced retinal neovascularization in OIR and several other mouse models (Campochiaro lab, JHU). Additionally, these effects are absent in control mice, where the transcriptome change is

minor and serves as quantitative evidence that corroborates the drug's high safety profile in clinical studies[14,17–19]. In fact, this type of four-arm design for transcriptome study is uncommon in the drug research field, but should be adapted more widely to identify both therapeutic effect and safety profile related to a drug.

2.6 Figures and Tables

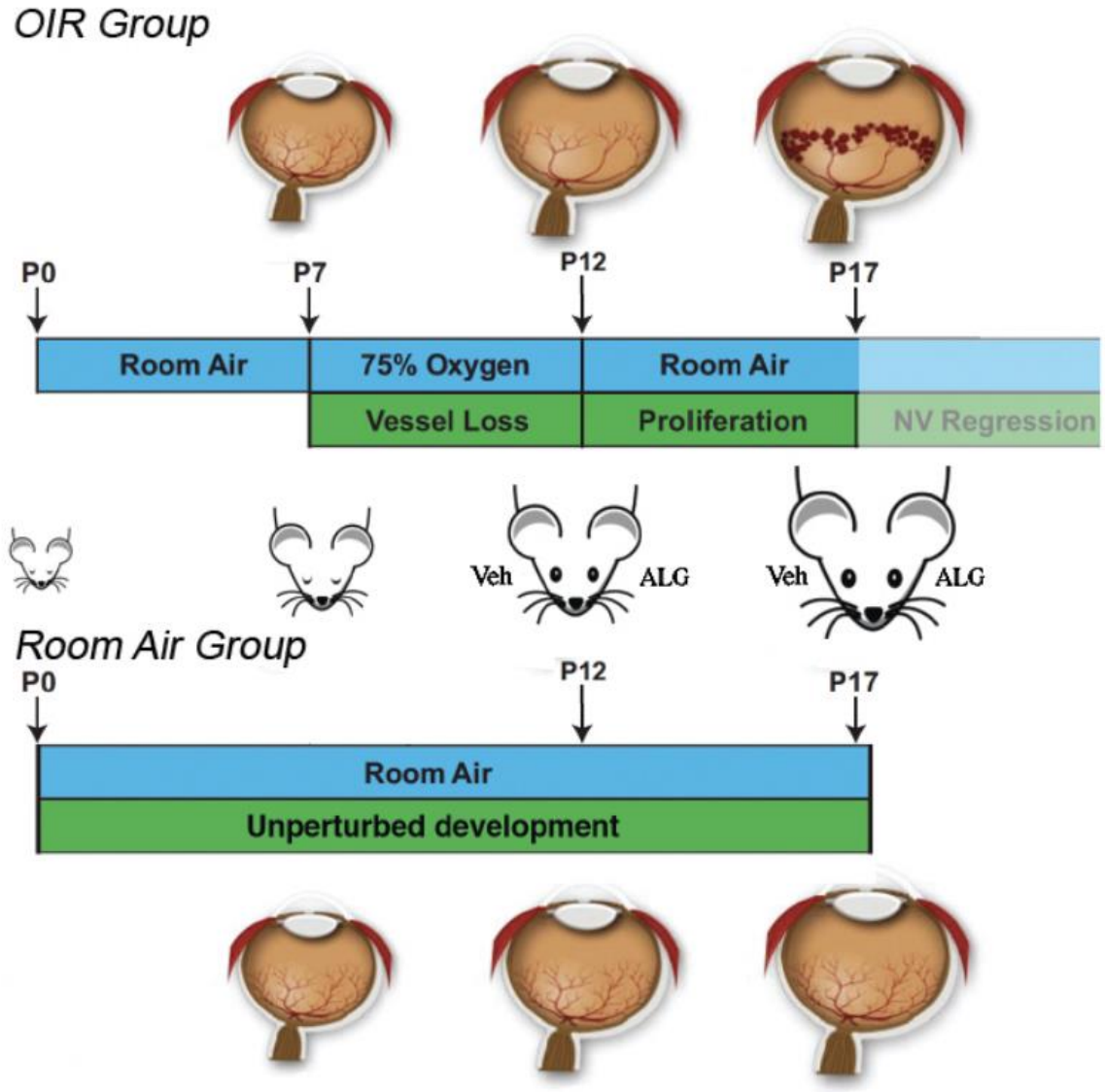

Figure 2.1 Generation of OIR and RA groups
Schematic showing treatments applied to all mice in the study. Mice were divided into two cohorts: OIR and RA (Room Air) groups. In the RA group, mice were left at room air for normal retinal development. In OIR group, mice were treated to 5 days of hyperoxia to regress retinal vessels, which then causes retinal angiogenesis that reaches peak development on P17. On P12, all mice are also injected intraocularly with either vehicle (Veh) or ALG-1001 (ALG). Retinal sample were collected on P17 for whole tissue RNA-seq. Schematic developed with significant modification from original figure by Stahl *et al.*[23].

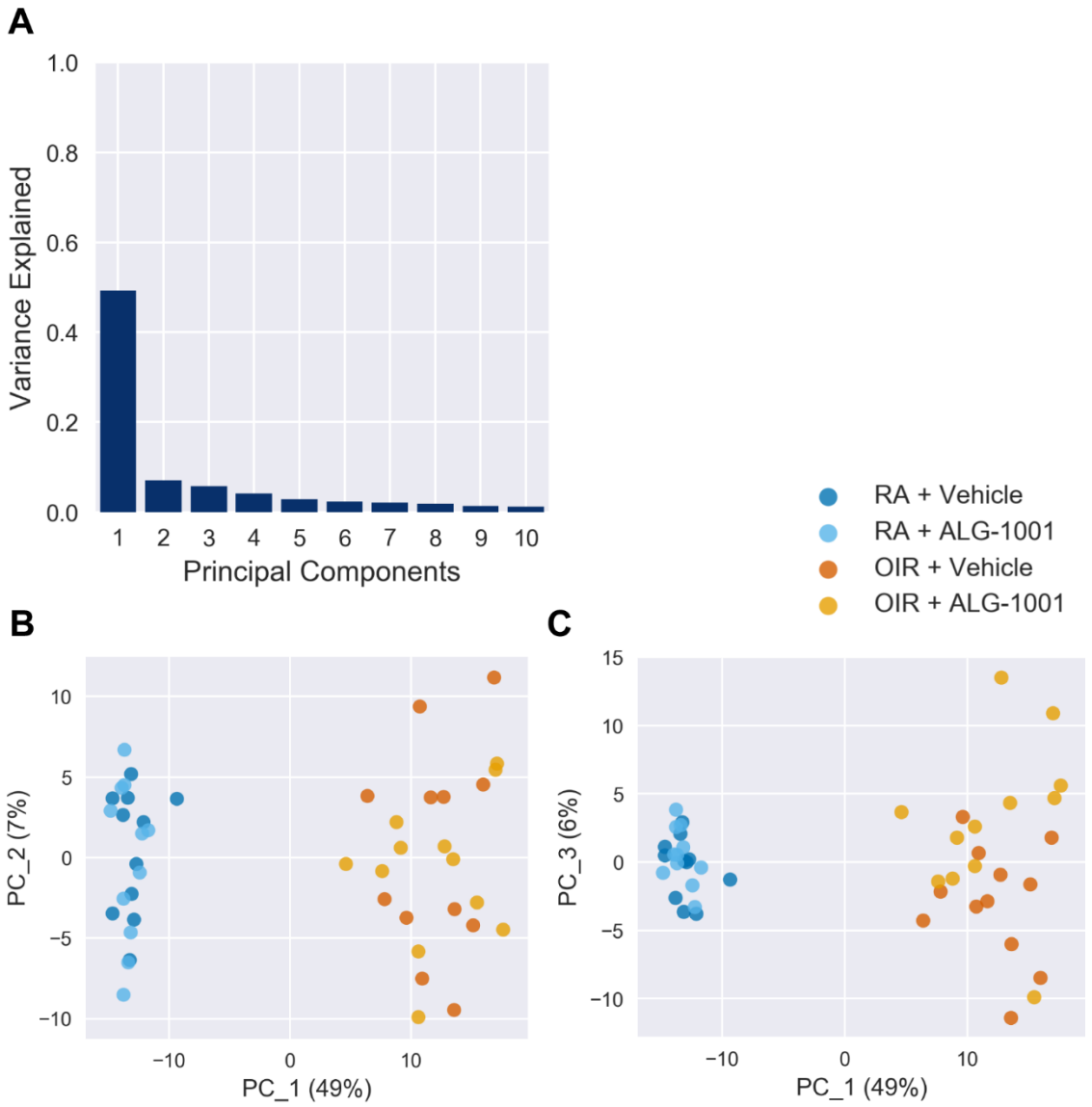

Figure 2.2 PCA with outlier pair of samples removed
$PCA_{No\text{-}outlier}$ with all samples except the outlier and its fellow eye sample. PC_# is principal component #. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC1 and PC2, PC1, and PC3, respectively. Percent of variance explained is shown in axis label. $PC1_{No\text{-}outlier}$ strongly separates OIR from RA samples.

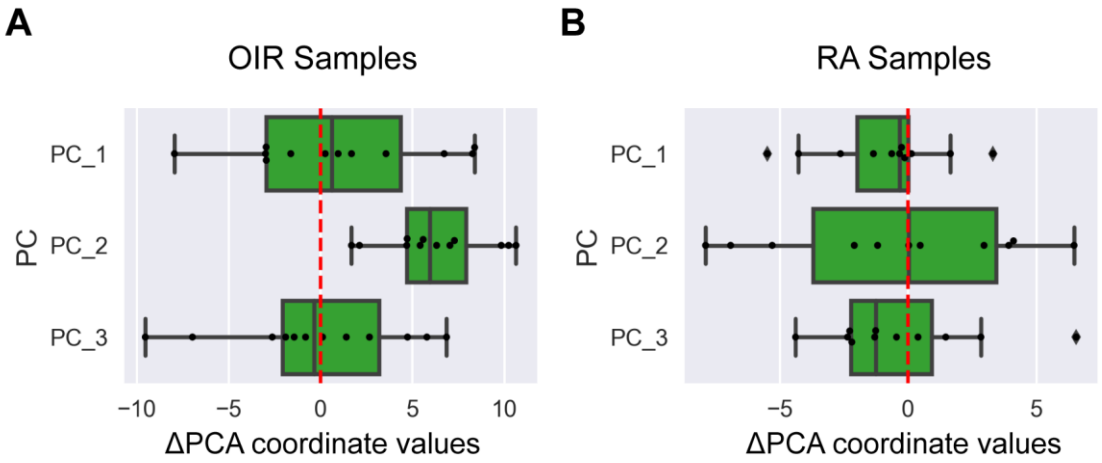

Figure 2.3 Animal-specific change in PCA coordinate values between ALG-1001 and vehicle treated eyes

PCA performed separately for OIR mice (**A**, PCA_{OIR} in **Figure S2.5**) and RA mice (**B**, PCA_{RA} in **Figure S2.6**). For each paired sample from the same animal, the difference in PCA transformed value is calculated and plotted. In OIR mice, ALG-1001 treated samples show a consistent shift in $PC2_{OIR}$ compared to fellow vehicle treated samples. Red line indicates zero difference. PC_# is principal component # and analysis excluded outlier sample pair.

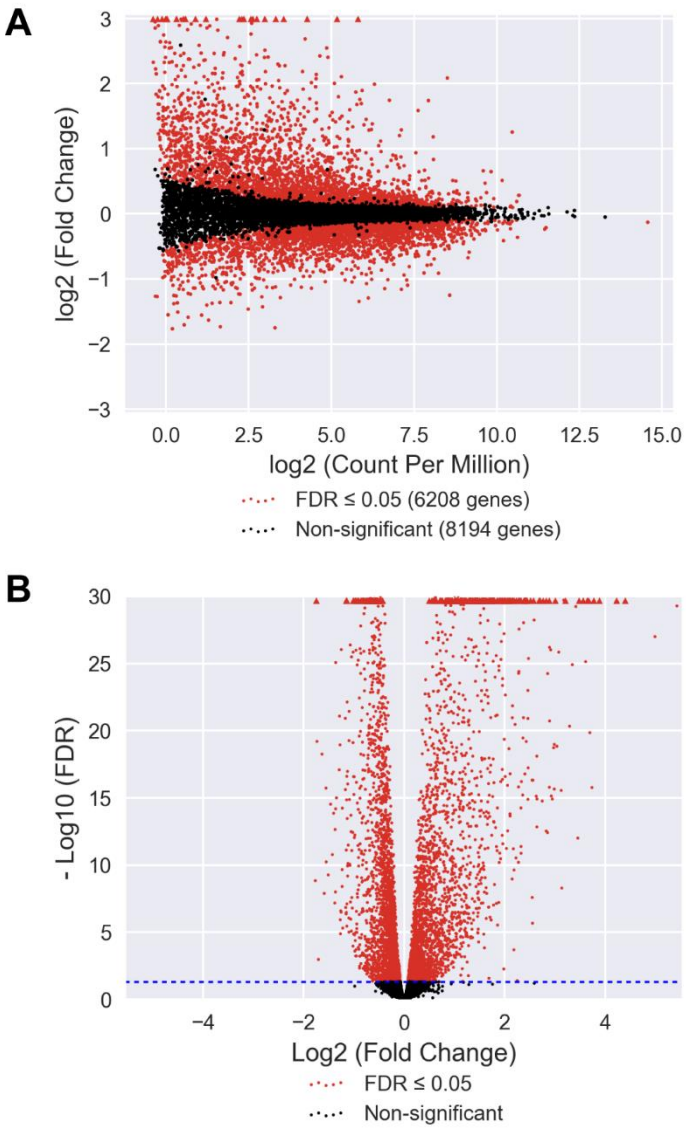

Figure 2.4 *edgeR* comparison that reveals hypoxia-modulated genes in retina
EdgeR comparison of OIR + vehicle relative to RA + vehicle as control. Modulated genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. **A.** Measured CPM and fold-change for each expressed gene. **B.** Measured fold-change and FDR displayed for each expressed gene. Blue dashed line indicates FDR=0.05.

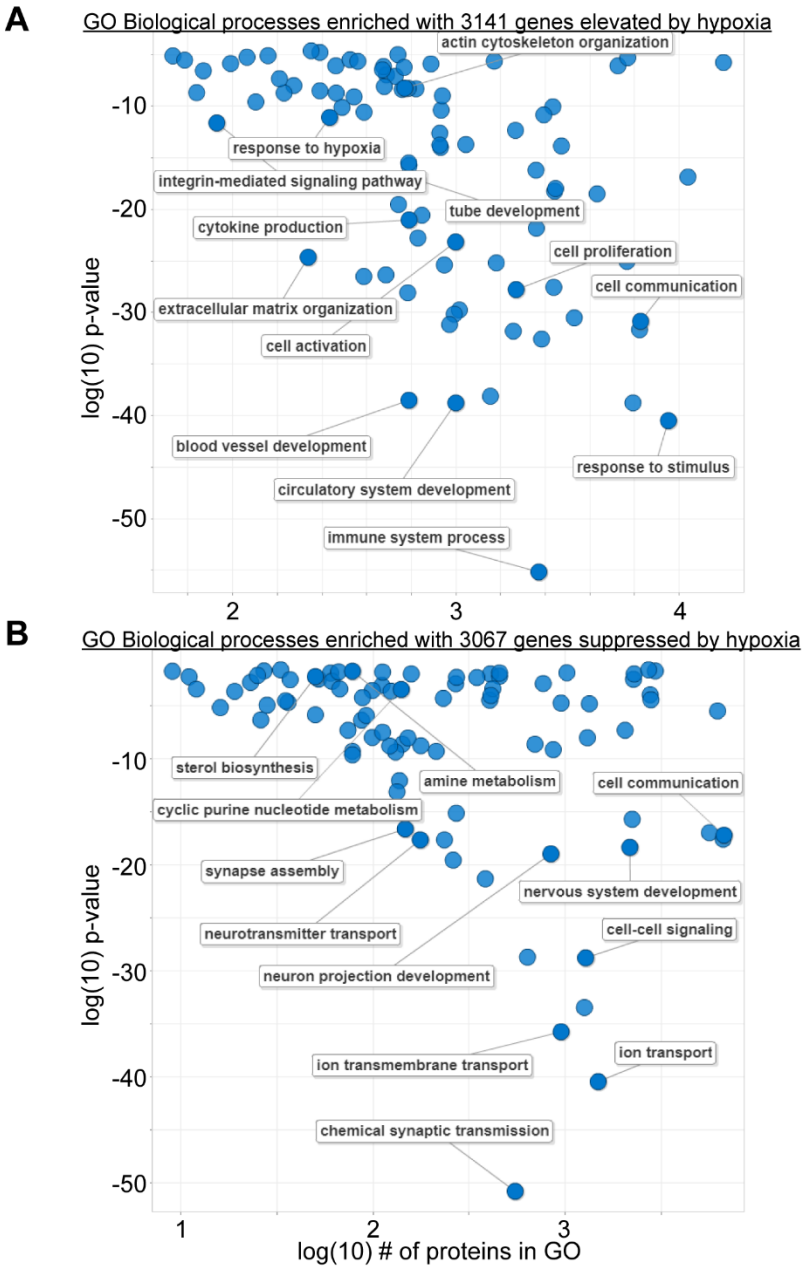

Figure 2.5 GO biological process enrichment with hypoxia-elevated and -suppressed genes

Hypoxia-elevated (**A**) and -suppressed (**B**) genes (OIR + Vehicle vs. RA + Vehicle) are submitted for GO enrichment with results visualized using *REVIGO*. In part **A**, many GO processes enriched are related to angiogenesis and inflammation. In part **B**, a significant number of GO processes enriched are related to neural system. P-value is *REVIGO*-derived GO enrichment p-value. Selected biologically relevant processes are labeled.

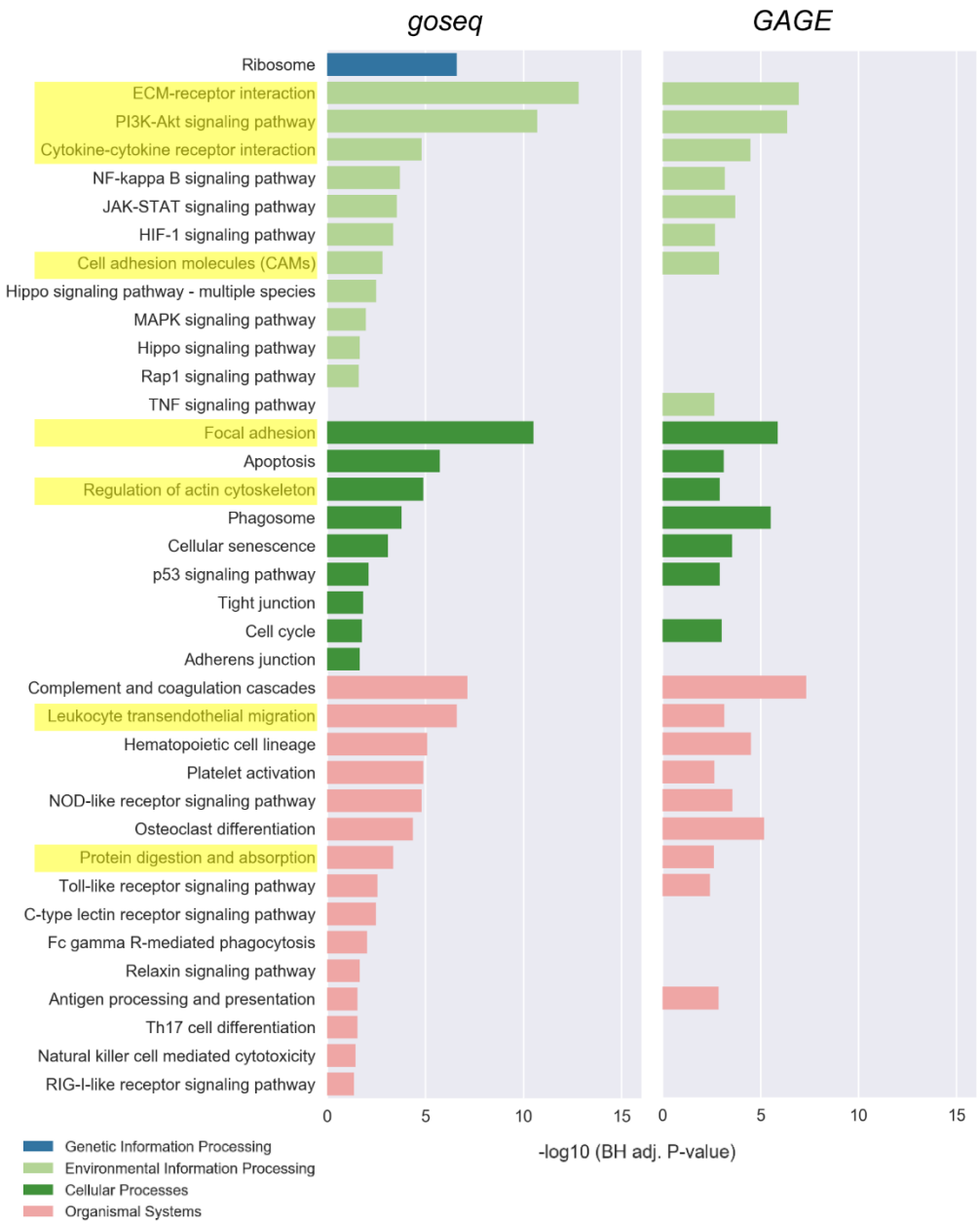

Figure 2.6 KEGG pathways enriched with hypoxia-elevated genes
Elevated KEGG pathways enriched with *goseq* using hypoxia-elevated genes (left) or *GAGE* with *edgeR* measured fold changes in comparison of OIR + Vehicle vs. RA + Vehicle (right). Many enriched pathways are related to inflammation, angiogenesis, cell proliferation and death. Highlighted pathways are enriched with ALG-1001-suppressed genes in OIR mice by *goseq* or *GAGE*. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis.

Figure 2.7 KEGG pathway enriched with hypoxia-suppressed genes
Suppressed KEGG pathways enriched with *goseq* using hypoxia-suppressed genes (left) or *GAGE* with *edgeR* measured fold changes in comparison of OIR + Vehicle vs. RA + Vehicle (right). Many enriched pathways are related to metabolism and neural system. No pathways overlap with those enriched with ALG-1001-elevated genes in OIR mice. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis.

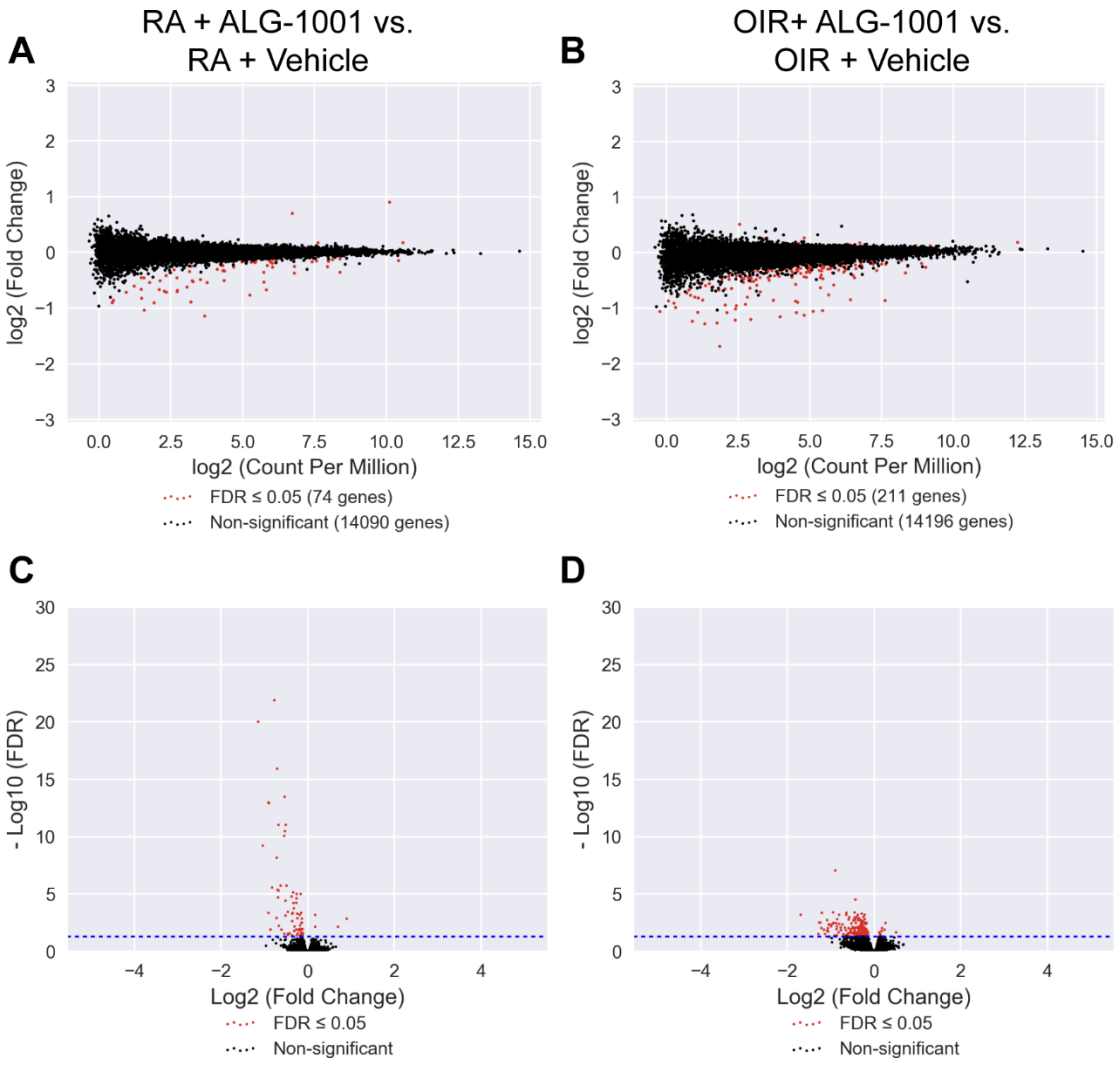

Figure 2.8 *edgeR* comparisons of ALG-1001 vs. vehicle treatment
EdgeR comparison of RA + ALG-1001 vs. RA + vehicle (**A, C**) and OIR + ALG-1001 vs. OIR + vehicle (**B, D**). Modulated genes are colored red while all others are black. **A-B.** Measured CPM and fold-change for each expressed gene. **C-D.** Measured fold-change and FDR displayed for each expressed gene. Blue dashed line indicates FDR=0.05.

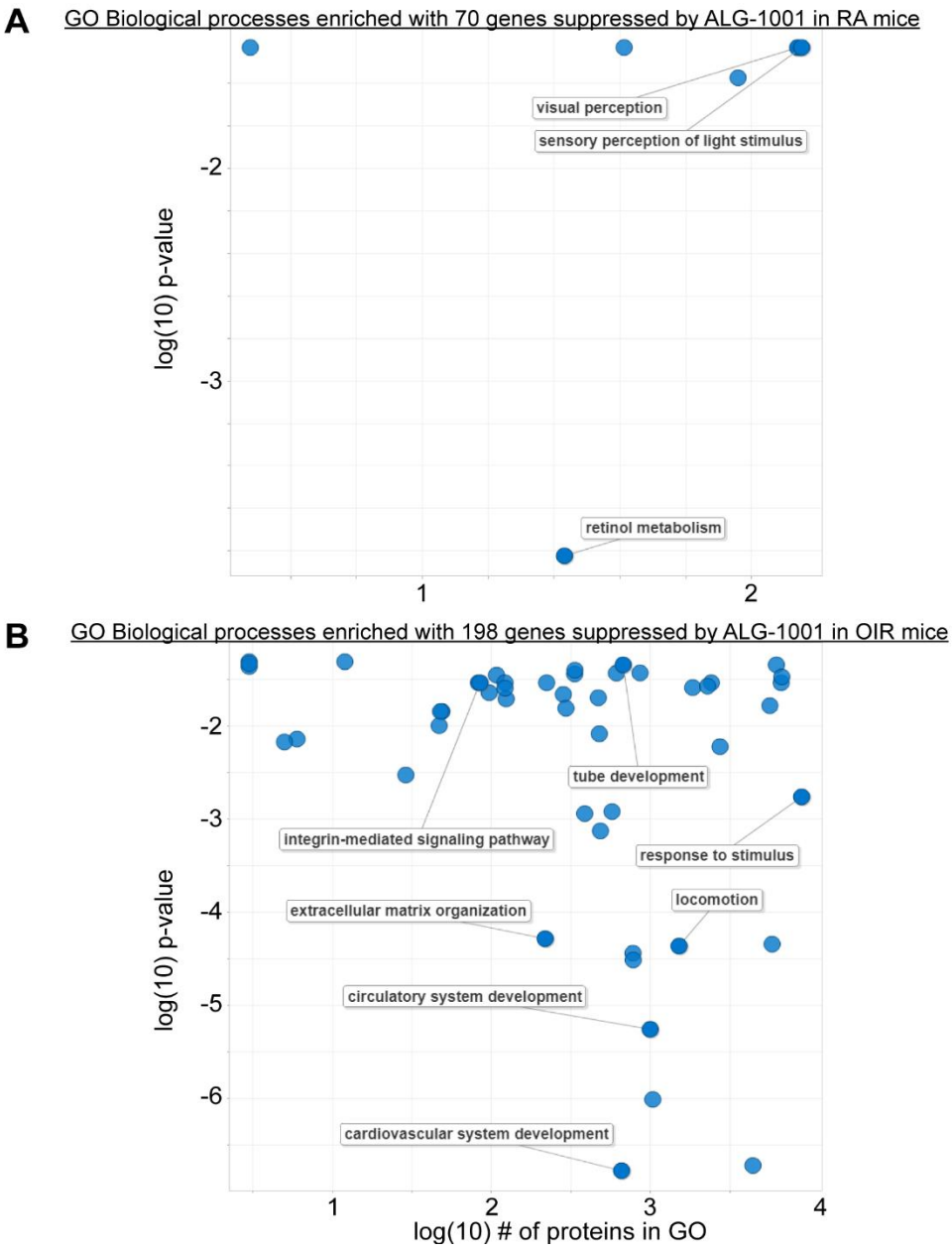

Figure 2.9 GO biological process enrichment of genes suppressed after ALG-1001 treatment

Suppressed genes after ALG-1001 treatment in RA mice (**A**) and OIR mice (**B**) relative to vehicle treated control are submitted for GO biological process enrichment with results visualized using *REVIGO*. Few processes with weak statistics are enriched with ALG-1001 treatment in RA mice, while those related to angiogenesis and inflammation are enriched with ALG-1001 treatment in OIR mice. P-value is *REVIGO*-derived GO enrichment p-value. Selected biologically relevant processes are labeled.

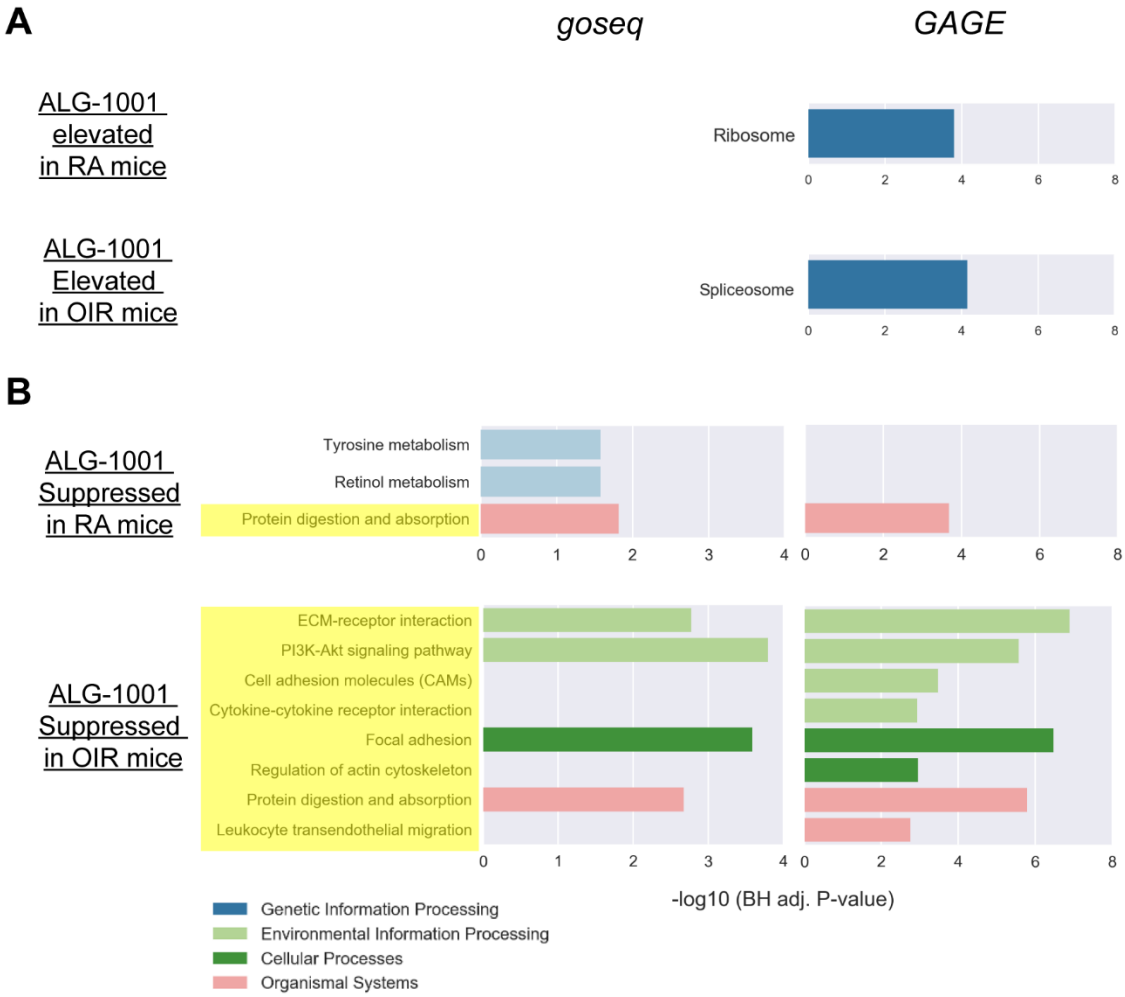

Figure 2.10 KEGG pathways enriched with ALG-1001-modulated genes
KEGG pathways significantly enriched with genes (**A**) elevated in the ALG-1001-treated eye relative to its fellow eye and (**B**) suppressed in ALG-1001-treated eye relative to its fellow eye. In each case, results for the RA mice are above those of the OIR mice. Results are shown for analysis with both *goseq* and *GAGE*. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis.

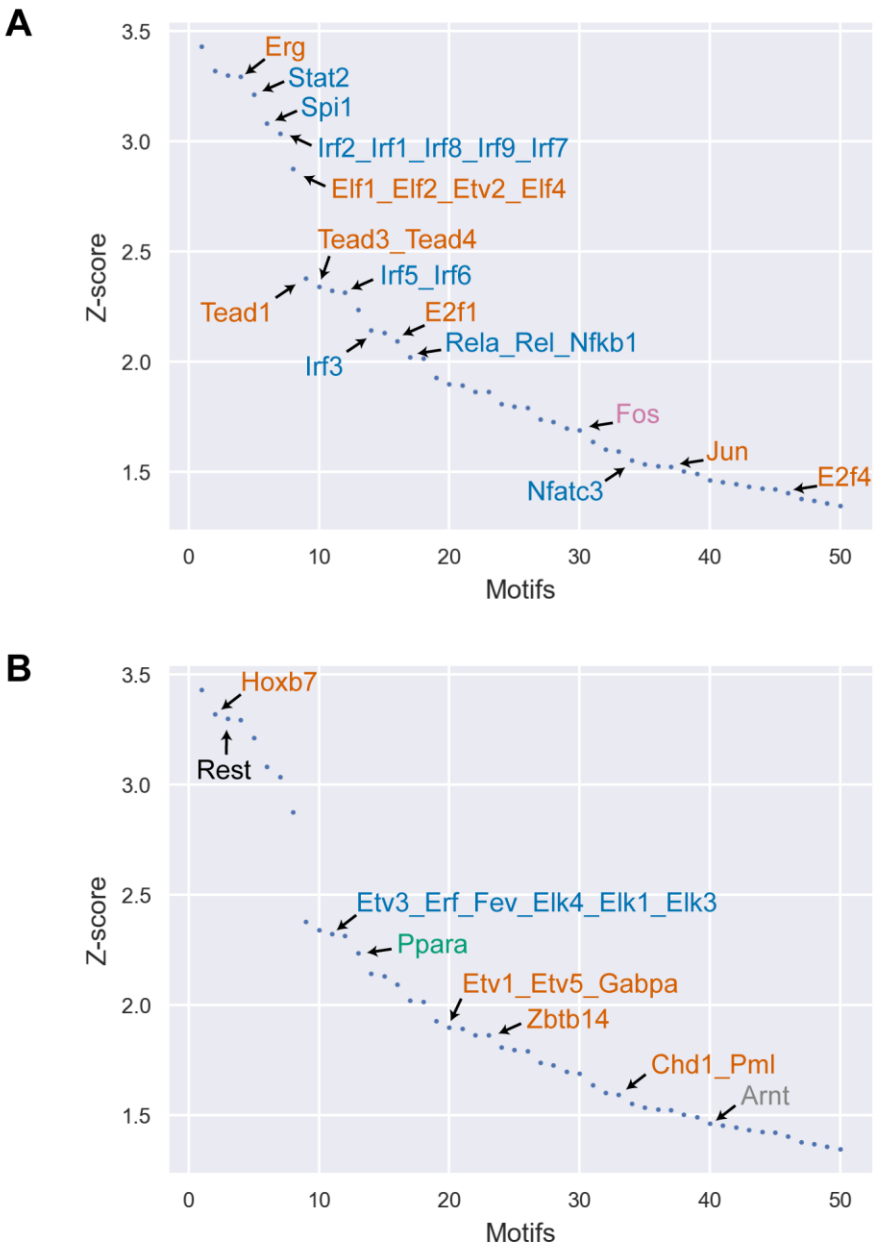

Figure 2.11 *ISMARA* enrichment for TFs associated with retinal hypoxia response
Raw reads of OIR + Vehicle and RA + Vehicle samples were submitted to *ISMARA* to enrich for TFs that are likely involved in regulating retinal hypoxia response. Plot shows the top 50 enriched TFs (**Table 2.2**) ranked by Z-score with: elevated (**A**) or suppressed (**B**) average target gene activity in OIR mice. Biologically relevant TFs are labeled with color by associated functions: cell proliferation and death (orange), angiogenesis (pink), inflammation (blue), neuronal development (black), metabolism (green), and hypoxia-response (gray).

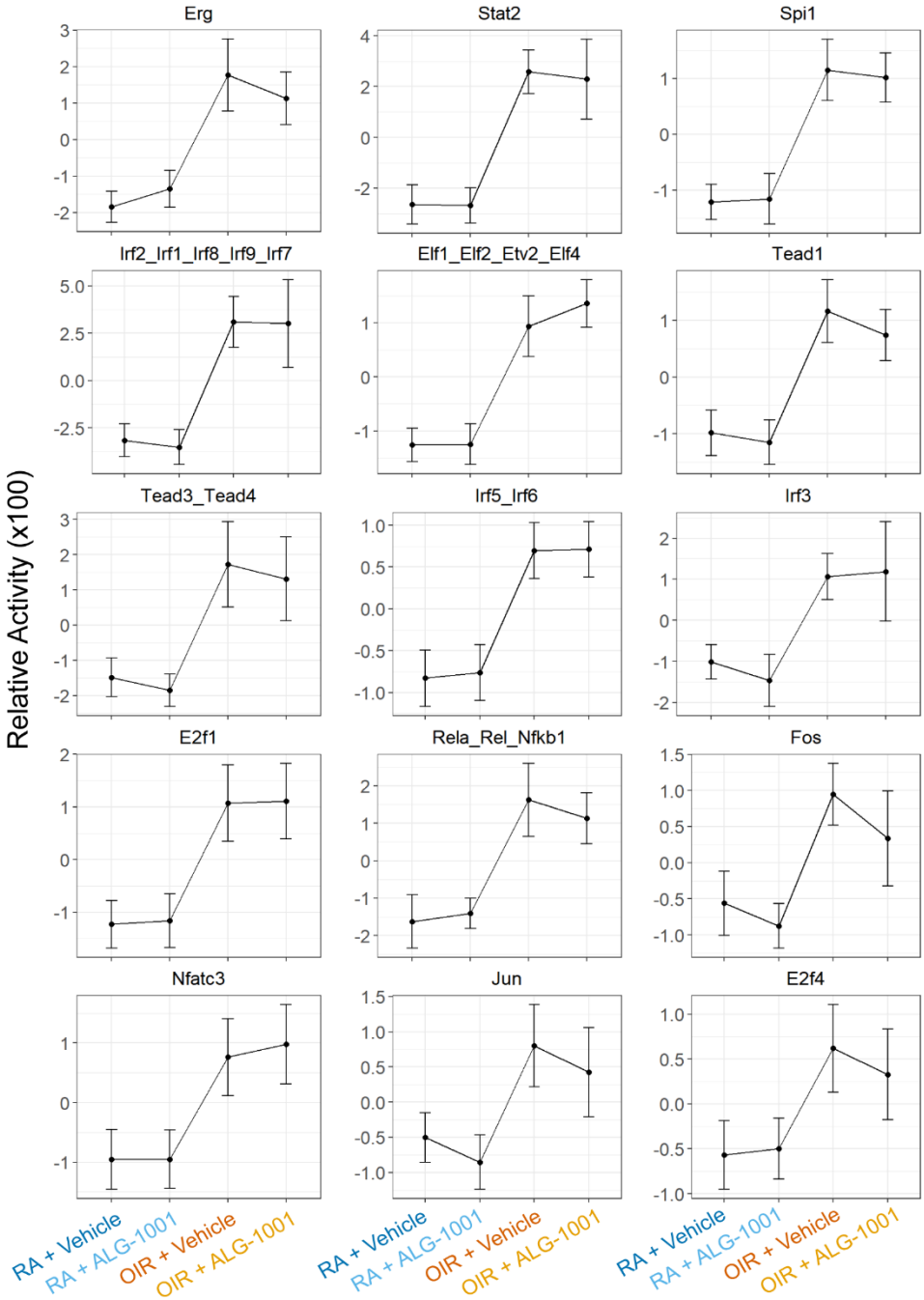

Figure 2.12 Activity profile of selected TF motifs with hypoxia-elevated target genes
OIR + Vehicle and RA + Vehicle samples were submitted to *ISMARA* to enrichment for TFs that are likely involved in retinal hypoxia transcriptome response. Activity profile (mean + standard deviation, values multiplied by 100) of selected TF motifs with elevated target activity in OIR vehicle condition are plotted.

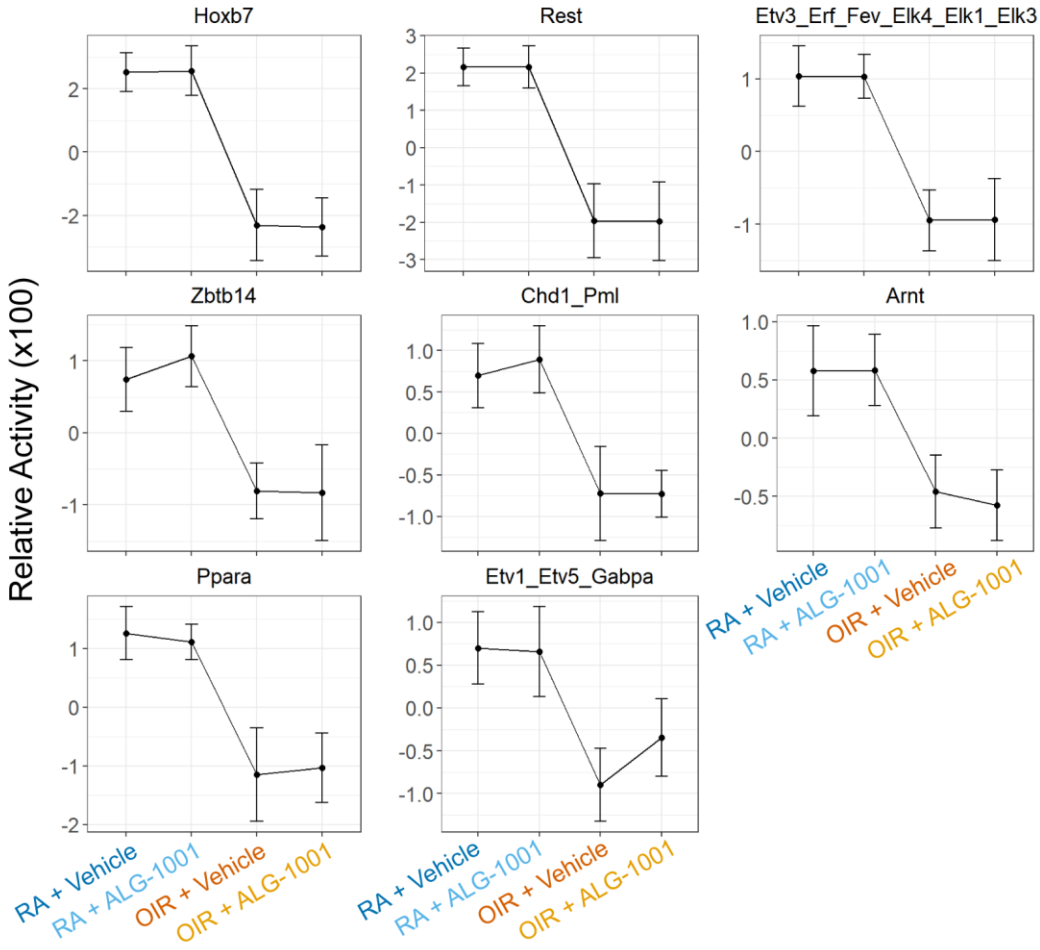

Figure 2.13 Activity profile of selected TF motifs with hypoxia-suppressed genes
OIR + Vehicle and RA + Vehicle samples were submitted to *ISMARA* to enrichment for TFs that are likely involved in retinal hypoxia transcriptome response. Activity profile (mean + standard deviation, values multiplied by 100) of selected TF motifs with suppressed target activity in OIR vehicle condition are plotted.

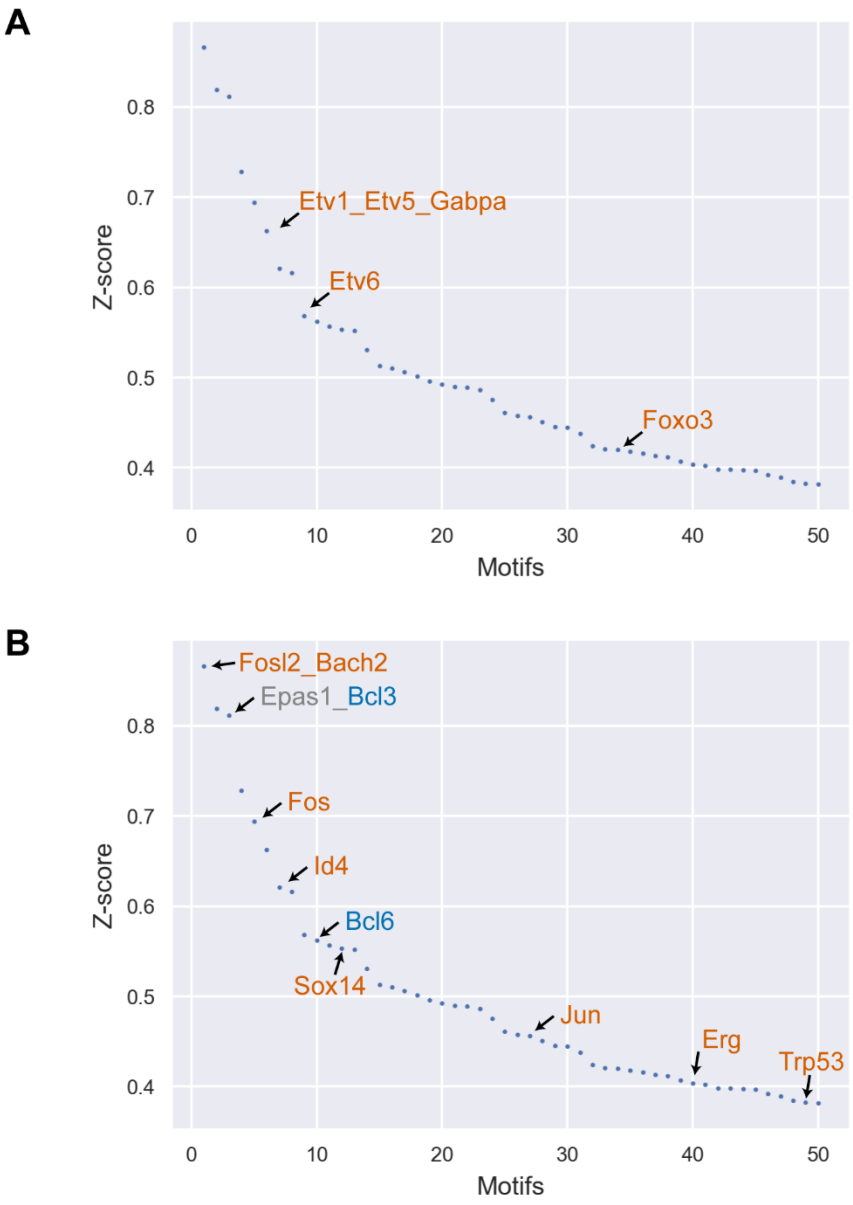

Figure 2.14 *ISMARA* enrichment for TFs associated with ALG-1001 treatment response in OIR mice

Raw reads of OIR + ALG-1001 and OIR + Vehicle samples were submitted to *ISMARA* to enrich for TFs that are likely involved in regulating ALG-1001's effect in OIR mice. Plot shows the top 50 enriched TFs (**Table 2.3**) ranked by Z-score with: elevated (**A**) or suppressed (**B**) average target gene activity in ALG-1001-treated samples. Biologically relevant TFs are labeled with color by associated functions: cell proliferation and death (orange), inflammation (blue), and hypoxia-response (gray).

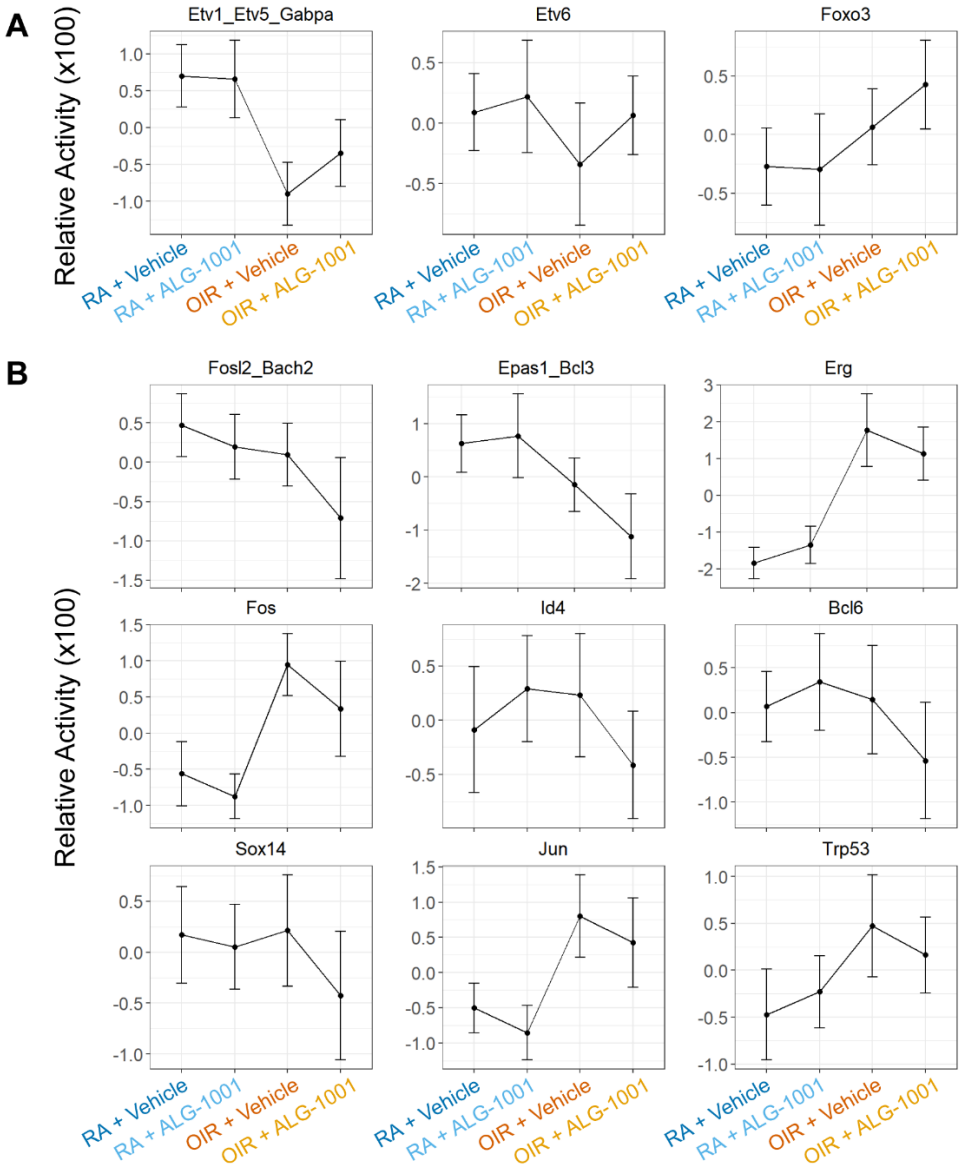

Figure 2.15 Activity profiles of selected TF motifs with ALG-1001-modulated genes in OIR

OIR + ALG-1001 and OIR + Vehicle samples were submitted to *ISMARA* to enrichment for TFs that are likely involved in regulating ALG-1001's effect in OIR mice. Activity profile (mean + standard deviation, values multiplied by 100) of selected TFs with (**A**) elevated or (**B**) suppressed target activity in OIR + ALG-1001 condition are plotted.

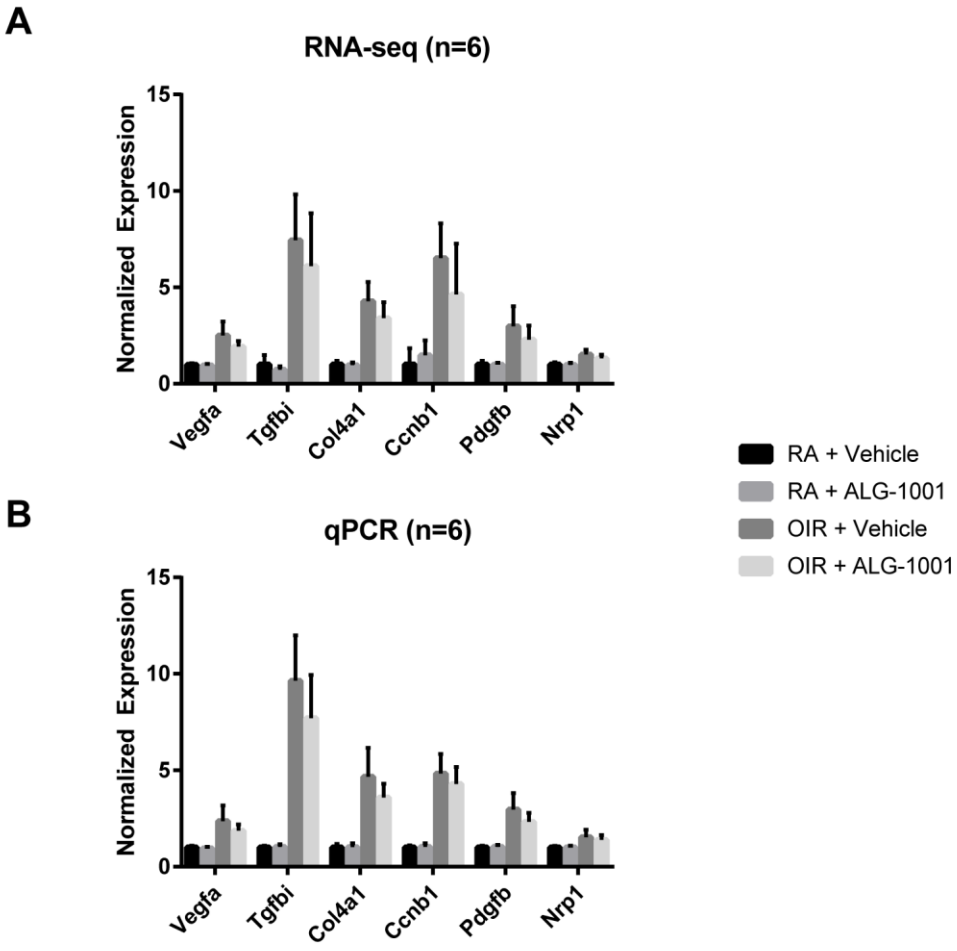

Figure 2.16 qPCR validation of selected RNA-seq results
Six genes shown by RNA-seq to be activated during retinal hypoxia (OIR + Vehicle vs. RA + Vehicle) and suppressed by ALG-1001 (OIR + ALG-1001 vs. OIR + Vehicle) are selected for qPCR validation. Quantified gene levels are normalized in both plots so that average RA + Vehicle level is 1. The mean and standard deviation are shown. **A.** RNA-seq gene level based on TPM values. **B.** qPCR gene level based on $\Delta\Delta C_t$-derived fold changes. RNA-seq and qPCR quantified gene levels are consistent across genes and conditions.

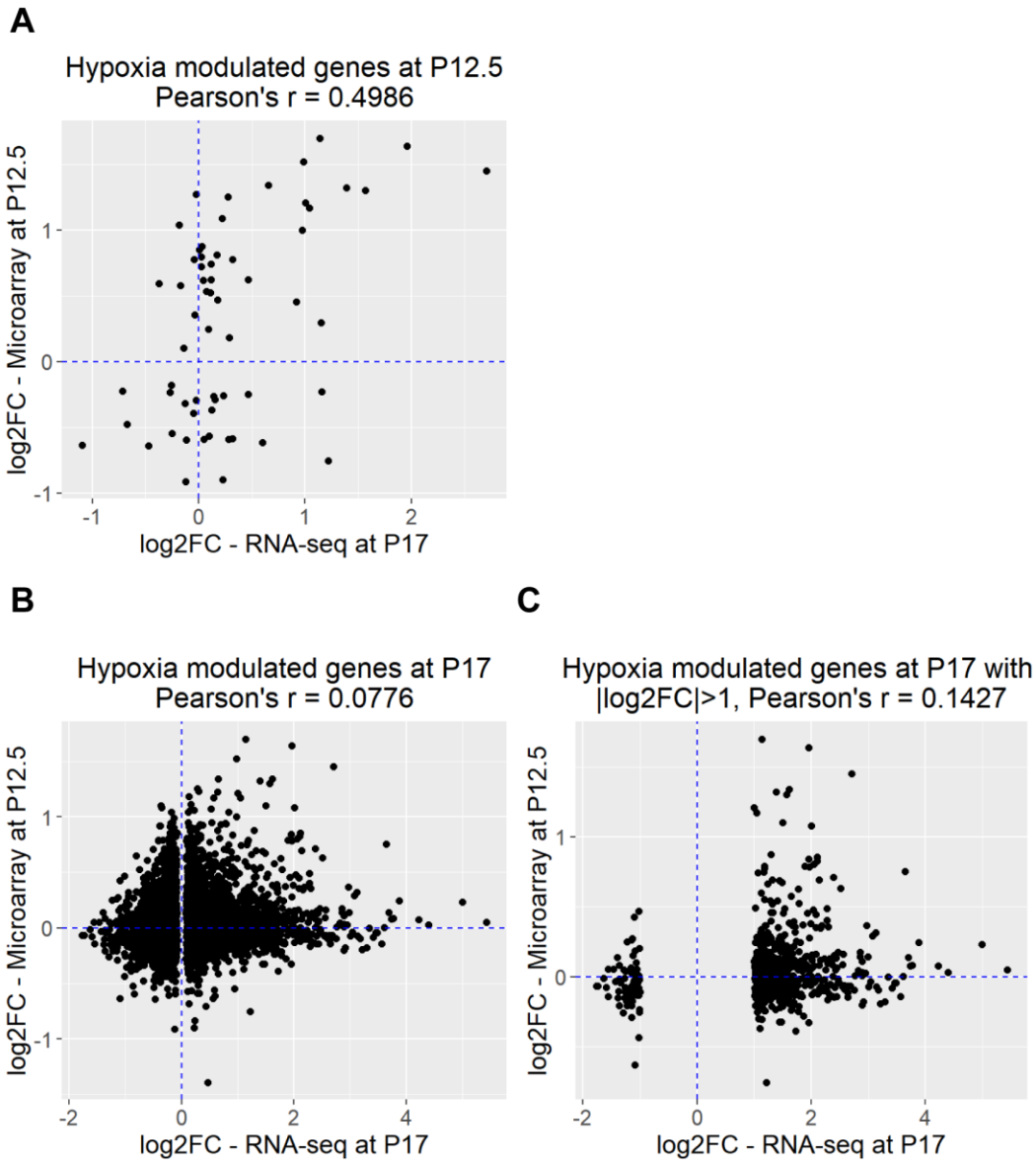

Figure 2.17 Compare P12.5 and P17 hypoxia-modulated genes
Analysis limited to hypoxia-modulated genes at (**A**) P12.5, (**B**) P17, and (**C**) P17 modulated genes with |log2-fold-change|>1 in the present study. X-axis shows fold changes in P17 mouse OIR retina compared to control (current study, vehicle comparison), Y-axis shows fold changes in P12.5 mouse OIR retina compared to control (Ishikawa *et al.*[58], untreated comparison). Dashed blue lines indicate Y=0 and X=0 lines. Log2FC is log2-fold-change. Pearson's correlation coefficients are shown in plot title. Correlation values show positive correlations across all gene sets tested, but significantly higher correlation when analysis is limited to hypoxia-modulated genes at P12.5.

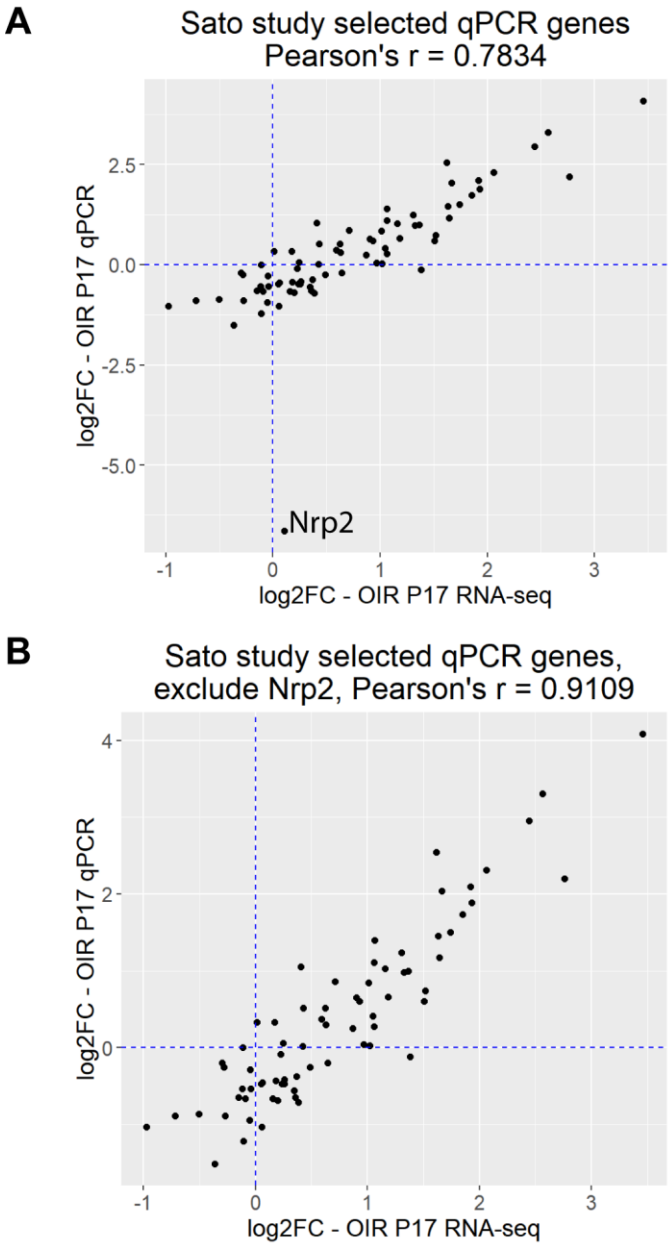

Figure 2.18 High correlation between qPCR from Sato study and RNA-seq
Genes in Sato study[97] are selected for visualization, with RNA-seq measured fold change (OIR + Vehicle vs. RA + Vehicle) on X-axis and qPCR (OIR vs. RA) measured fold change on Y-axis. Dashed blue lines indicate Y=0 and X=0 lines. Pearson's correlation coefficients are shown in plot title. All genes measured in the Sato study that are also measurable with RNA-seq are plotted in (**A**). Nrp2 appears to be an outlier and is removed in (**B**) with Pearson's correlation re-calculated. With Nrp2 removed, a relatively high correlation is observed between the two datasets.

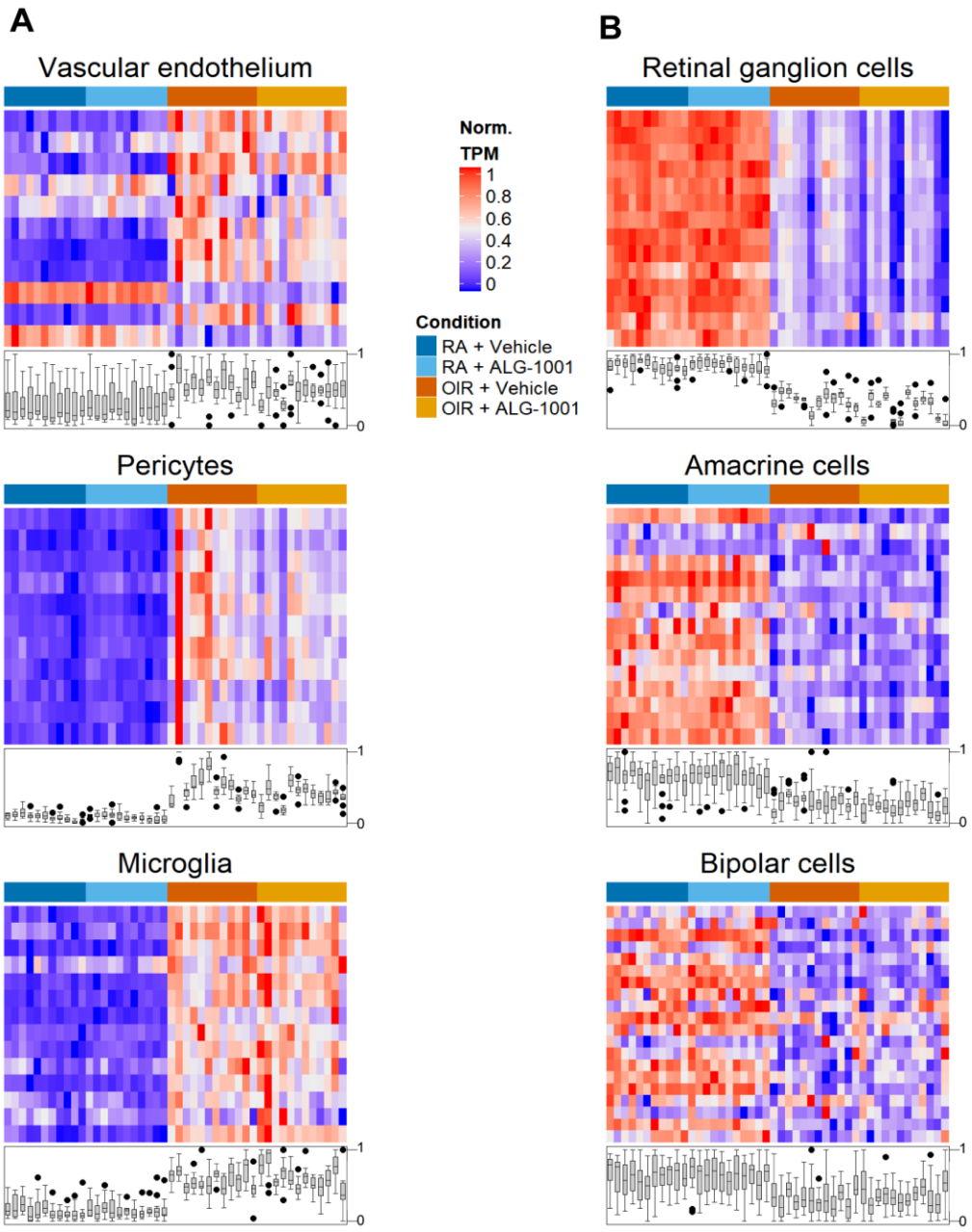

Figure 2.19 Retinal cell type-specific gene level change heatmap – part 1
Normalized TPM (Norm. TPM) values of retinal cell type-specific genes are visualized in form of heatmap. All non-outlier samples are sorted column-wise by condition as indicated by color bar on top of heatmap. Norm. TPM values of each sample are summarized by box plot below heatmap. Cell types are categorized by >80% (**A**) or <20% (**B**) of markers with positive hypoxia-induced fold change (OIR + Vehicle vs. RA + Vehicle), suggesting evidence for specific cell type expansion or depletion after hypoxia.

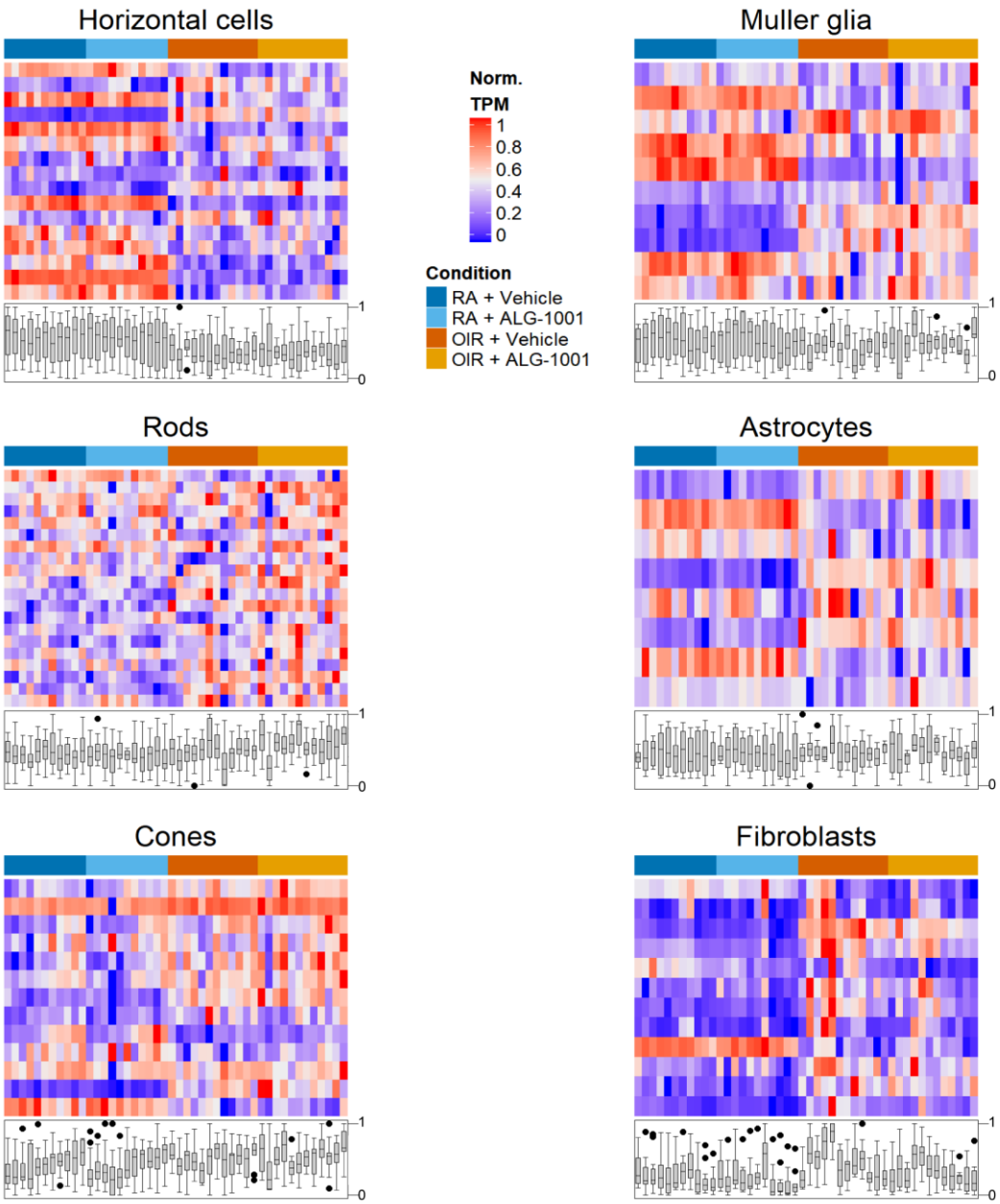

Figure 2.20 Retinal cell type-specific gene level change heatmap – part 2
Normalized TPM (Norm. TPM) values of retinal cell type-specific genes are visualized in form of heatmap. All non-outlier samples are sorted column-wise by condition as indicated by color bar on top of heatmap. Norm. TPM values of each sample are summarized by box plot below heatmap. All cell types have within 20-80% of markers with positive hypoxia-induced fold change (OIR + Vehicle vs. RA + Vehicle), suggest lack of strong evidence for cell composition change by hypoxia.

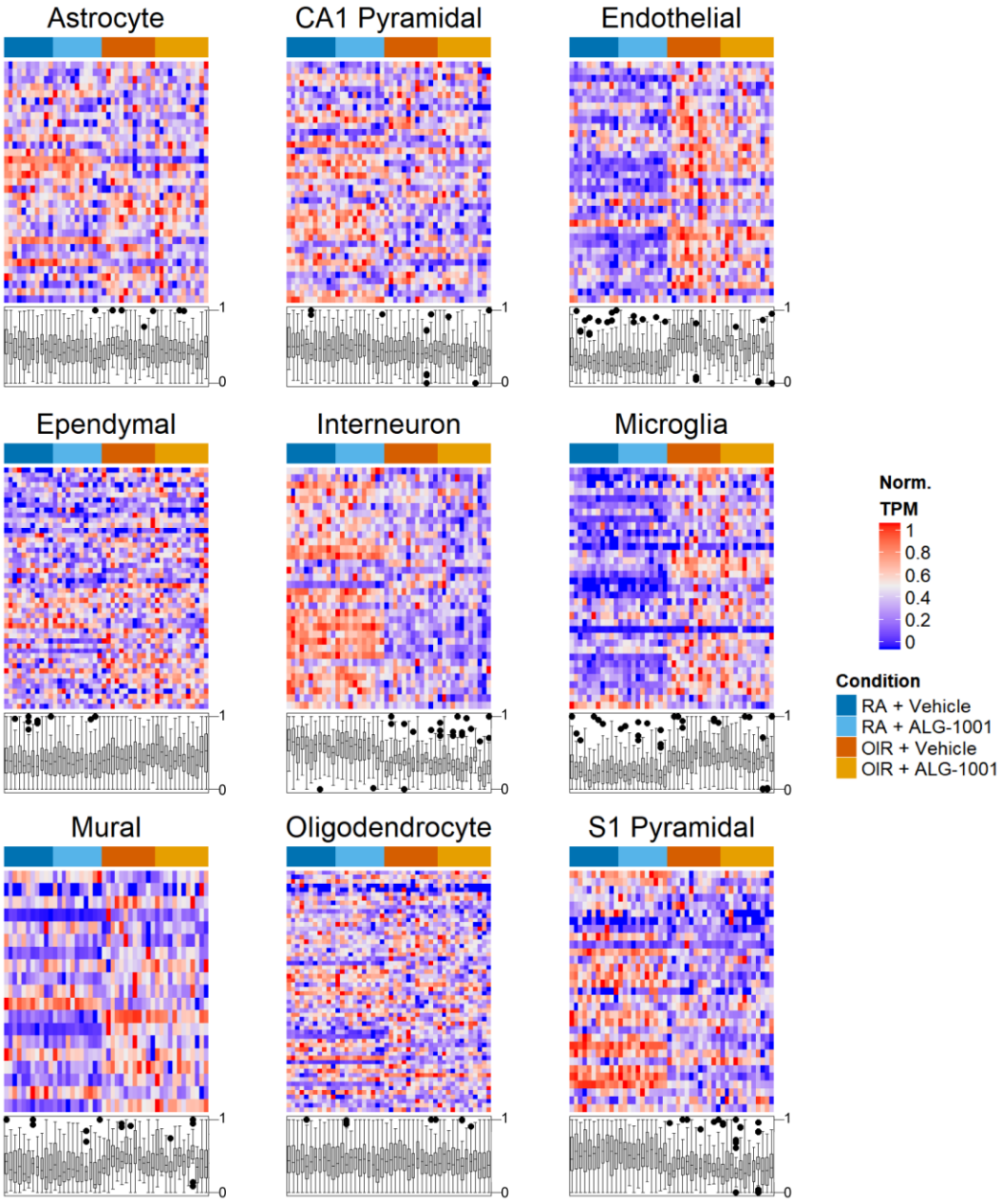

Figure 2.21 Brain cell type-specific gene level heatmap

Normalized TPM (Norm. TPM) values of brain cell type-specific genes are visualized in form of heatmap. All non-outlier samples are sorted column-wise by condition as indicated by color bar on top of heatmap. Norm. TPM values of each sample plotted as a heatmap and summarized by box plot below heatmap. Endothelial and microglia cells show transcriptome sign of cell number expansion after hypoxia while interneuron show sign of cell number depletion.

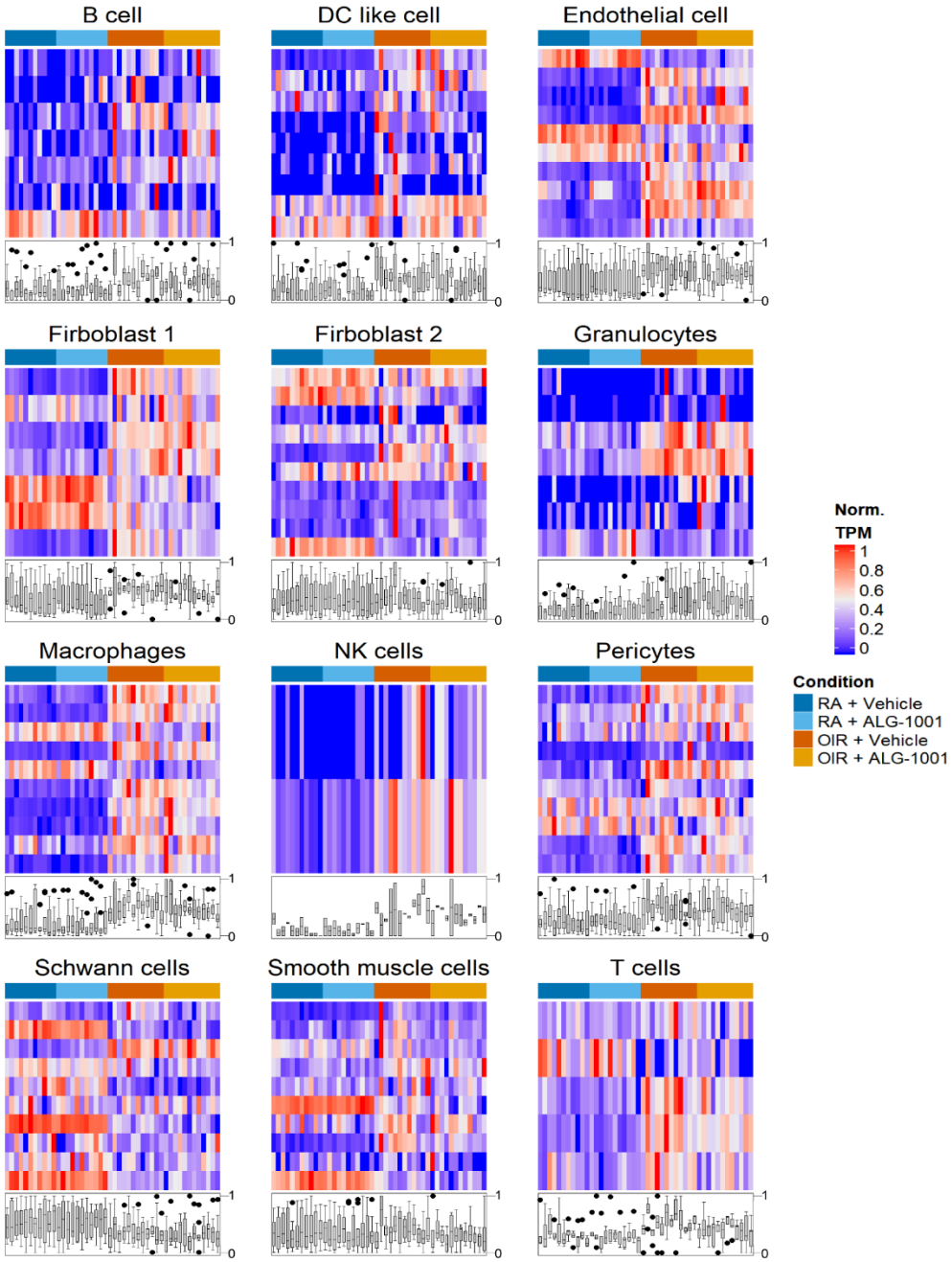

Figure 2.22 Heart and immune cell type-specific gene level heatmap

Normalized TPM (Norm. TPM) values of heart and immune cell type-specific genes are visualized in form of heatmap. All non-outlier samples are sorted column-wise by condition as indicated by color bar on top of heatmap. Norm. TPM values of each sample plotted as a heatmap and summarized by box plot below heatmap. Endothelial, macrophage and pericytes cells show transcriptome signs of cell number expansion after hypoxia.

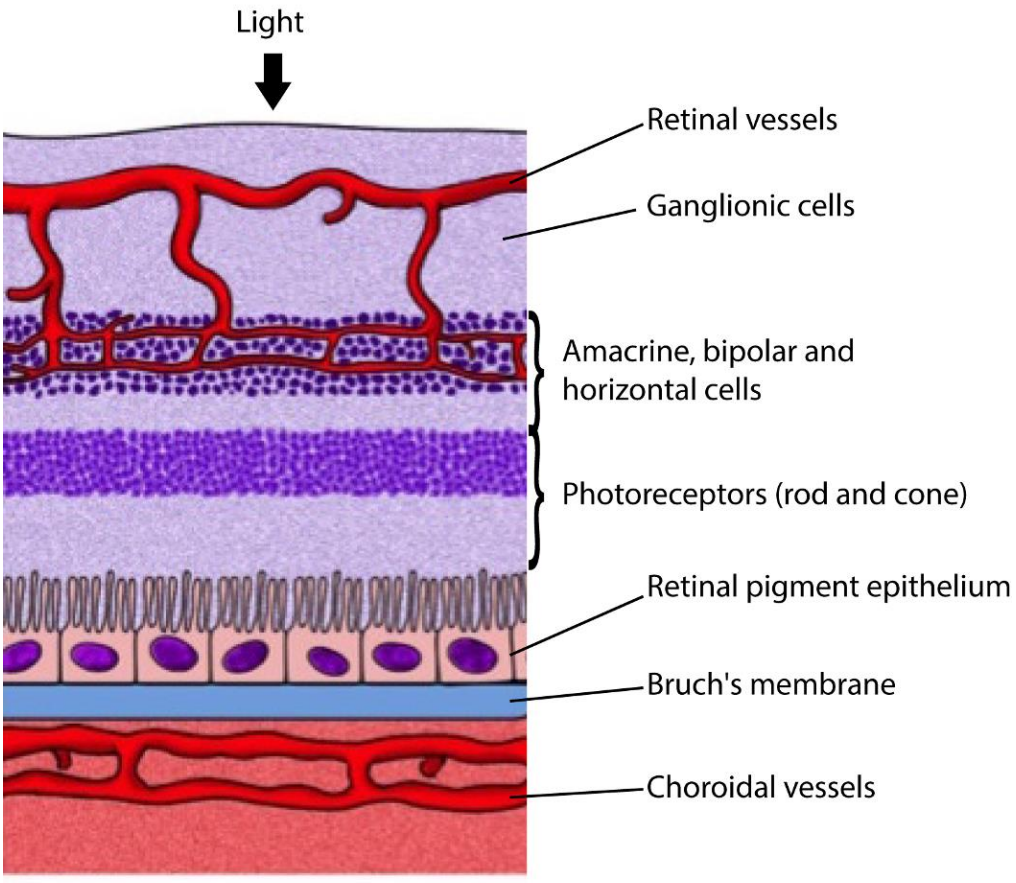

Figure 2.23 Blood vessels that support the retina
The retinal tissue is composed of distinguishable cell layers and is supported by two systems of blood vessel: choroidal and retinal vessels. In diabetic retinopathy and related diseases, neovascularization originates from the retinal vessels. In wet age-related macular degeneration, vessel growth stems from chorodial vessel and expand into the retina through Bruch's membrane and retinal pigment epithelium.

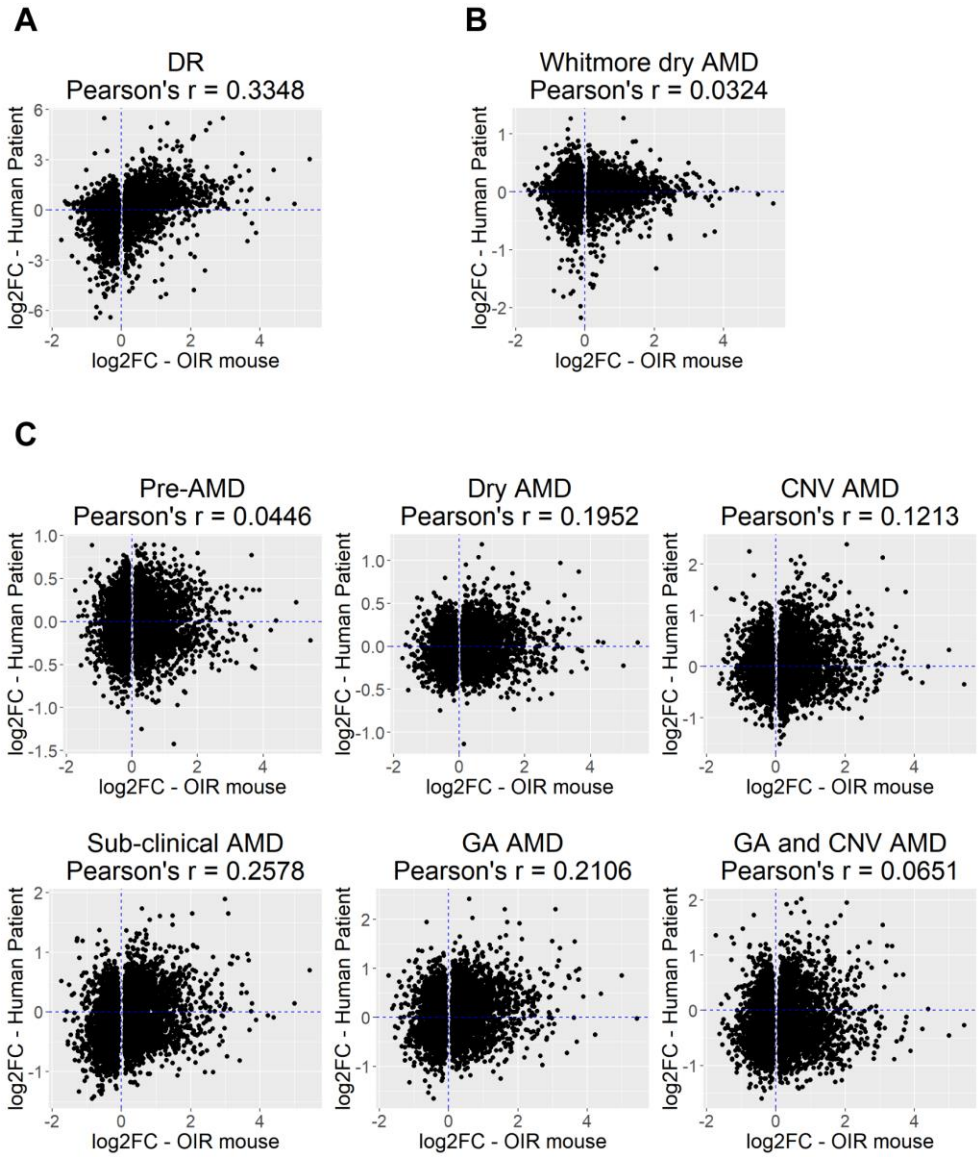

Figure 2.24 Hypoxia-modulated genes in OIR show positive correlation with DR and AMD

Hypoxia-modulated genes in OIR + Vehicle vs. RA + Vehicle were selected for plotting with their log2-fold-change on the X-axis. Y-axis shows log2-fold-change derived from human clinical studies of DR (**A**, Ishikawa study[56]) and AMD (**B**, Whitmore study[62] and **C**, Newman study with samples from various stages of AMD[61]) patients versus healthy cohorts. Pearson's correlations are calculated for each plot and displayed in the title. Correlation to Whitmore study is low due to different tissues tested (RPE and choroid tissue in Whitmore study versus retina in the present study). Positive correlations are observed across all comparisons with the highest between OIR mice and DR patients.

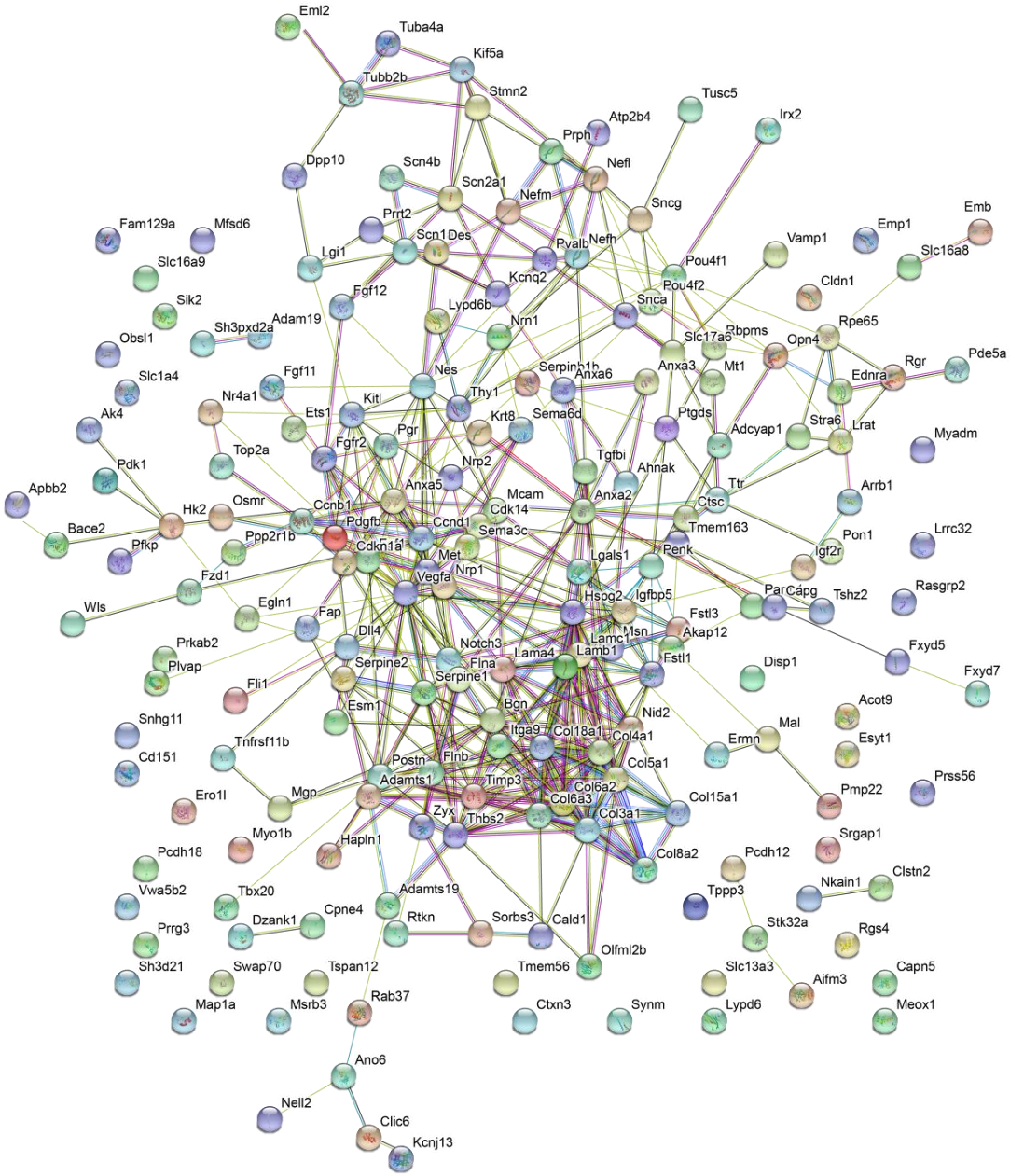

Figure 2.25 Protein-protein interaction graph of ALG-1001-suppressed genes in OIR mice

Suppressed genes in the comparison OIR + ALG-1001 vs. OIR + Vehicle were submitted to *STRING* to visualize connection between the proteins these genes encode. Closely-associated proteins were clustered into two observable clusters that include proteins that regulate angiogenesis, cell growth, motility and adhesion.

Gene	Direction	5' to 3' Sequence
Gapdh	Forward	AGGTCGGTGTGAACGGATTTG
Gapdh	Reverse	TGTAGACCATGTAGTTGAGGTCA
Vegfa	Forward	GCACATAGAGAGAATGAGCTTCC
Vegfa	Reverse	CTCCGCTCTGAACAAGGCT
Tgfbi	Forward	CAGCACGGCCCCAATGTAT
Tgfbi	Reverse	GGGACCTTTTCATATCCAGGACA
Col4a1	Forward	CTGGCACAAAAGGGACGAG
Col4a1	Reverse	ACGTGGCCGAGAATTTCACC
Ccnb1	Forward	AGAGCTATCCTCATTGACTGGC
Ccnb1	Reverse	AACATGGCCGTTACACCGAC
Pdgfb	Forward	TGCTGCACAGAGACTCCGTA
Pdgfb	Reverse	GATGAGCTTTCCAACTCGACTC
Nrp1	Forward	GACAAATGTGGCGGGACCATA
Nrp1	Reverse	TGGATTAGCCATTCACACTTCTC

Table 2.1 qPCR primers

Transcription Factor	Z-score	Activity diff. OIR + Vehicle – RA + Vehicle	Activity diff. OIR + ALG-1001 – OIR + Vehicle
Esrrb_Esrra	3.430265	-0.03231	-0.00108
Hoxb7	3.320527	-0.05074	-0.00091
Rest	3.299146	-0.04594	-0.00013
Erg	3.294661	0.038089	-0.00408
Stat2	3.211164	0.055767	-0.00141
Spi1	3.082239	0.024399	-0.00136
Irf2_Irf1_Irf8_Irf9_Irf7	3.033734	0.064852	-0.0005
Elf1_Elf2_Etv2_Elf4	2.875717	0.022166	0.004149
Tead1	2.379707	0.023405	-0.00212
Tead3_Tead4	2.340649	0.034684	-0.00265
Etv3_Erf_Fev_Elk4_Elk1_Elk3	2.322505	-0.02009	0.001095
Irf5_Irf6	2.314453	0.016595	0.000054
Ppara	2.237315	-0.0259	0.000813
Irf3	2.143088	0.022035	0.000424
Spic	2.130711	0.031155	0.00171
E2f1	2.093629	0.023454	0.000368
Rela_Rel_Nfkb1	2.019875	0.035114	-0.00309
Nhlh1	2.013827	-0.01724	0.000541
Hsfy2	1.929333	0.018104	-0.00114
Etv1_Etv5_Gabpa	1.897489	-0.01673	0.004982
Nfil3_Tef	1.893919	-0.02395	0.000293
Runx2_Bcl11a	1.864103	0.015979	0.000558
Zbtb14	1.86305	-0.01603	-0.00019
Onecut1_Cux2	1.810252	-0.01471	-0.00137
Spib	1.796229	0.022922	0.001562
Tcf7l1	1.791399	0.020335	-0.00218
Nkx3-2	1.739507	0.018187	-0.00146
Vdr	1.726826	-0.01608	0.001313
Klf4_Sp3	1.699623	0.01458	-0.00276
Fos	1.691026	0.016206	-0.00398
Srf	1.638139	0.023866	-0.00307
Nfic_Nfib	1.602884	0.02684	-0.00318
Chd1_Pml	1.592378	-0.01424	-5.6E-05

Nfatc3	1.554134	0.018816	0.000899
Cebpb	1.53586	0.023171	-0.00088
Myog_Tcf12	1.527287	-0.01169	-0.00254
Jun	1.522605	0.014055	-0.00274
Brca1	1.502649	-0.01066	-0.00058
Nfe2_Bach1_Mafk	1.493496	-0.01506	-0.00204
Arnt	1.462555	-0.01142	-0.00049
Nr2c2	1.454343	0.020481	0.002312
Foxf1	1.445879	-0.01458	-0.00062
Nrf1	1.433137	-0.01262	0.006486
Hey2	1.426067	-0.00987	-0.0005
Hoxa4	1.421377	-0.01097	0.001539
E2f4	1.404575	0.013226	-0.00129
E2f2_E2f5	1.380132	0.021169	-0.00248
Nr2e1	1.370329	-0.00996	0.00128
Snai1_Zeb1_Snai2	1.358701	-0.01225	0.002979
Pou2f2_Pou3f1	1.346542	0.019262	0.000696

Table 2.2 Top 50 TF motifs enriched with hypoxia

TF motifs are ordered by Z-score calculated by *ISMARA* between OIR + Vehicle and RA + Vehicle. "Activity diff. OIR + Vehicle – RA + Vehicle" is the average target gene activity difference between OIR + Vehicle and RA + Vehicle conditions. "Activity diff. OIR + ALG-1001 – OIR + Vehicle" is the average target gene activity difference between OIR + ALG-1001 and OIR + Vehicle conditions. Average target gene activity value was calculated by *ISMARA* and is a measure of average expression of genes that contain the TF motif[47]. A positive activity difference indicates higher average expression of target genes in the first condition of the comparison. Bolded TFs indicate those that are also found in the list of top 50 TF motifs identified with ALG-1001 treatment in hypoxia condition. Selected biologically relevant TFs are labeled in **Figure 2.11** and their activity profile in **Figure 2.12 and 2.13**.

Transcription Factor	Z-score	Activity diff. OIR + Vehicle – RA + Vehicle	Activity diff. OIR + ALG-1001 – OIR + Vehicle
Fosl2_Bach2	0.866357	-0.0049	-0.00509
Zfp263	0.819355	0.012223	-0.00343
Epas1_Bcl3	0.811259	-0.00834	-0.00519
Zfx_Zfp711	0.728028	-0.00205	0.004629
Fos	0.693695	0.016206	-0.00398
Etv1_Etv5_Gabpa	0.662583	-0.01673	0.004982
Id4	0.620802	0.003379	-0.00362
Zfp652	0.616116	0.011292	-0.00288
Etv6	0.567969	-0.005	0.002375
Bcl6	0.562167	0.000266	-0.00336
Nr1i2	0.556931	0.005574	-0.00342
Sox14	0.552916	0.000295	-0.00319
Nfe2l2	0.552207	-0.00472	-0.00257
Tcf21_Msc	0.530699	-0.00055	0.002127
Klf1	0.512749	0.00123	-0.00515
Nrf1	0.510102	-0.01262	0.006486
Nr5a2	0.506052	0.000782	0.002753
Pitx3	0.501015	0.012145	0.00327
Hes5_Hes7	0.496144	-0.00357	0.001916
Tfap2d	0.492704	0.007685	0.002097
Hmx1	0.489585	0.005463	0.002019
Elf1_Elf2_Etv2_Elf4	0.489097	0.022166	0.004149
Rfx5	0.486523	0.011514	0.002021
Zfp219_Zfp740	0.475588	0.000603	0.003176
Hcfc1_Six5_Smarcc2_Zfp143	0.461353	-0.00939	0.005259
Atf3	0.457336	-0.00924	0.003532
Jun	0.456271	0.014055	-0.00274
Myf6	0.451113	-0.0023	-0.00301
Zic4	0.445612	-0.00533	0.001703
Bsx	0.444895	0.004805	-0.00175
Tead1	0.437919	0.023405	-0.00212
Rorc_Nr1d1	0.424381	0.001066	0.001643
Cux1	0.42068	-0.00216	0.002518

Foxo3	0.420198	0.003627	0.00162
Zbtb4	0.41824	-0.00849	0.001714
Srf	0.416023	0.023866	-0.00307
Taf1	0.412891	-0.01311	0.007399
Mafb	0.411795	0.001309	0.001874
Arntl_Tfe3_Mlx_ Mitf_Mlxipl_Tfec	0.406988	0.005772	0.002271
Erg	0.40352	0.038089	-0.00408
Tcf7l1	0.401934	0.020335	-0.00218
Pitx2_Otx2	0.398509	0.01238	0.003785
Myog_Tcf12	0.398378	-0.01169	-0.00254
Gsx1_Alx1_Mixl1_Lbx2	0.397332	-0.00475	-0.00163
Vsx1_Uncx_Prrx2_ Shox2_Noto	0.396899	-0.00391	-0.00281
Elf5	0.392256	-0.00897	0.003303
Gsc2_Dmbx1	0.389319	-0.00164	0.001911
Sox6_Sox9	0.384688	0.015104	-0.0018
Trp53	0.382392	0.010513	-0.00165
Nfe2_Bach1_Mafk	0.381502	-0.01506	-0.00204

Table 2.3 Top 50 TF groups identified with ALG-1001 treatment under hypoxia
TF motifs are ordered by Z-score calculated by *ISMARA* between OIR + ALG-1001 and OIR + Vehicle. "Activity diff. OIR + Vehicle – RA + Vehicle" is the average target gene activity difference between OIR + Vehicle and RA + Vehicle conditions. "Activity diff. OIR + ALG-1001 – OIR + Vehicle" is the average target gene activity difference between OIR + ALG-1001 and OIR + Vehicle conditions. Average target gene activity value was calculated by *ISMARA* and is a measure of average expression of genes that contain the TF motif[47]. A positive activity difference indicates higher average expression of target genes in the first condition of the comparison. Bolded TFs indicate those that are also found in the list of top 50 TF motifs enriched with hypoxia exposure. Selected biologically relevant TFs are labeled in **Figure 2.14** and their activity profile in **Figure 2.15**.

Cell type	Genes
Horizontal cells	Calb1;Tpm3;Sept4;Slc4a3;Vim;Gnas;Pcsk1n;Smarca4;Snhg11; Ndrg1;Snap25;Cd47;Maged1;Tfap2b;Mgarp;Ndrg4;Ppp1r1a
Retinal ganglion cells	Nefl;Nefm;Stmn2;Sncg;Uchl1;Stmn3;Thy1;Slc17a6;Nrn1;Gap43; Cdk14;Ywhah;Tubb2a;Calm2
Amacrine cells	Ppp1r17;Tcf4;Meg3;Syt1;Ebf3;Pnmal2;Marcks;Lgr5;Celf4;Elavl3; Gria2;Ebf1;Basp1;Snca;Sphkap
Rods	Rho;Sag;Pdc;Gngt1;Gnat1;Prph2;Gnb1;Pde6g;Pde6b;Rcvrn;Rp1; Tulp1;Rom1;Cnga1;Nr2e3;Nrl;Slc24a1;Hmgn1;Rpgrip1;Rs1
Cones	Pde6h;Opn1sw;Gngt2;Opn1mw;Arr3;Gnat2;Pde6c;Kcne2;Guca1a; Cd59a;Ccdc136;Gnb3;Scg3
Bipolar cells	Trpm1;Pcp2;Calm1;Gng13;Isl1;Gnao1;Pcp4;Prkca;Nme1;Car8; Chgb;Gpr179;Vsx2;Lrtm1;Qpct;Lin7a;Ablim1;Grm6;Map4;Gm4792
Muller glia	Acsl3;Rlbp1;Slc1a3;Car2;Spc25;Car14;Mfge8;Crym;Rtn4; Gpr37
Astrocytes	Igf2;Aldoc;S100b;Pdgfra;Gpm6b;Id3;Cntnap2;Mlc1
Fibroblasts	Crhbp;Optc;Col18a1;Gja1;Ptgds;Fbn2;Fstl1;Fbln1;Tsc22d1;Pvrl3; Col23a1;Timp3
Vascular endothelium	Bsg;Flt1;Cldn5;Ptprb;Itm2a;Ramp2;Cd93;Fn1;Spock2;Ctla2a;Pltp
Pericytes	Rgs5;Mgp;Cald1;Aspn;Nid1;Itga1;Serpine2;Pdgfrb;Myl9;Itgb1; Higd1b
Microglia	Ctss;Hexb;C1qb;Lgmn;C1qc;C1qa;B2m;Sepp1;Cx3cr1;Ly86; Laptm5;Jun;Ctsb;Csf1r

Table 2.4 Mouse retina cell type-specific marker genes
The set of cell type-specific markers are derived from the Macosko *et al.* scRNA-seq dataset[50]. The selected marker genes are found to be expressed at a high level in each cell type and not in the other cell types included in this analysis.

Human Disease Comparison	Modulated genes in OIR[1]	Non-modulated genes in OIR[2]	Modulated gene in OIR with GO term[3]				
			Immune system process	Angiogenesis	Cell death	Metabolic process	Nervous system development
Diabetic Retinopathy (Ishikawa)							
DR	0.3347	0.059	0.3948	0.4073	0.3484	0.3293	0.2999
Age-related Macular Degeneration (Newman)							
Pre-AMD	0.0459	0.0108	-0.0042	0.0641	0.0631	0.0411	0.1096
Sub-clinical AMD	0.2606	0.0386	0.2613	0.1367	0.2767	0.2561	0.3292
Dry AMD	0.199	0.052	0.2307	0.1765	0.2592	0.2189	0.2761
GA AMD	0.216	0.0284	0.2818	0.2264	0.2627	0.2355	0.2477
CNV AMD	0.123	0.0043	0.2179	0.0477	0.1751	0.1243	0.1638
GA and CNV AMD	0.0654	-0.0205	0.1194	-0.0415	0.0625	0.0602	0.0898
Age-related Macular Degeneration (Whitmore)							
Dry AMD	0.0324	0.0116	0.0286	0.1102	0.0339	0.0075	0.018

Table 2.5 Pearson's correlation of fold changes between OIR and disease studies

[1]All hypoxia-modulated genes found in OIR + Vehicle vs. RA + Vehicle comparison are included for calculation.

[2]All non-hypoxia-modulated genes found expressed in mouse retina are included for calculation.

[3]Pearson's r calculated with genes included in the following GO biological processes: "immune system process", "angiogenesis", "cell death", "metabolic process" and "nervous system development". Aside from "metabolic process", all others are enriched in *goseq* enrichment analysis. "Metabolic process" is selected as it encompasses several related metabolic processes that are enriched in *goseq* analysis.

Correlation to Whitmore study is low due to different tissues analyzed (RPE and choroid tissue in Whitmore study versus retina in Ishikawa, Newman, and the present studies).

Study	Method	Sample size	OIR time point	Study of hypoxia	Study of drug	Raw data available	Modulated gene list available
Hoppe[141]	RNA-seq	6	P12	No	Yes	No	No
Ishikawa[58]	Microarray	3	P12.5	Yes	No	Yes	Yes
Yang[142]	Microarray	3	P13	Yes	No	No	Yes
Oubaha[143]	Microarray	2	P14	Yes	No	No	No
Sato[97]	qPCR, 94 genes	1	P12-P21	Yes	No	Yes	No

Table 2.6 Mouse OIR transcriptome studies

Study	Method	Sample size	Disease	Raw data available	Modulated gene list available
Ishikawa[56]	Microarray	3 control, 6 DR	DR	Yes	Yes
Lam[110]	RNA-seq	4 control, 8 DR	DR	No	No
Newman[61]	Microarray	31 control, 37 AMD	AMD	Yes	Yes
Hunter[144]	Microarray	11 control, 10 AMD	AMD	No	Yes, but discovery with mixed samples
Radeke[6]	Microarray	10 control, 5 AMD	AMD	No	Yes, but between macular and extramacular samples
Whitmore[62]	Microarray	9 control, 9 AMD	AMD	Yes	No

Table 2.7 Human retinal disease transcriptome studies
Comparative analysis limited to Ishikawa, Lam, Newman, and Whitmore studies due to availability of raw data (Ishikawa, Newman, and Whitmore) and GO biological processes enriched using genes with statistically significant change (Lam). Hunter study is excluded due to both a lack of raw data and mixing of samples from different AMD stages and tissue specimens. Radeke study is excluded as it did not study expression difference between AMD and control cohorts.

Gene Ontology	**Selected Gene Frequency**	**Total Frequency**	**Percent of Total**
Biological process	61	20952	0.3
Developmental process	61	5717	1.1
Anatomical structure development	61	5271	1.2
Multicellular organism development	61	4846	1.3
System development	61	4354	1.4
Anatomical structure morphogenesis	*61*	*2409*	*2.5*
Anatomical structure formation involved in morphogenesis	*61*	*1050*	*5.8*
Circulatory system development	*61*	*1030*	*5.9*
Cardiovascular system development	*61*	*699*	*8.7*
Vasculature development	*61*	*687*	*8.9*
Blood vessel development	*61*	*663*	*9.2*
Blood vessel morphogenesis	**61**	**568**	**10.7**
Angiogenesis	**61**	**476**	**12.8**
Cellular process	59	16081	0.4
Regulation of cellular process	55	11255	0.5
Response to stimulus	52	9183	0.6
Regulation of developmental process	*45*	*2380*	*1.9*
Localization	44	5544	0.8
Cell communication	43	6871	0.6
Signaling	43	6807	0.6
Cell differentiation	42	3789	1.1
Metabolic process	41	11008	0.4
Organic substance metabolic process	41	10448	0.4
Nitrogen compound metabolic process	41	9491	0.4
Signal transduction	41	6394	0.6
Cellular metabolic process	38	9910	0.4
Regulation of metabolic process	35	6173	0.6

Regulation of cellular metabolic process	35	5833	0.6
Regulation of nitrogen compound metabolic process	35	5613	0.6
Response to stress	35	3398	1
Locomotion	*35*	*1544*	*2.3*
Regulation of vasculature development	**35**	**310**	**11.3**
Regulation of angiogenesis	**35**	**283**	**12.4**
Negative regulation of cellular process	34	4487	0.8
Animal organ development	34	3217	1.1
Cell motility	*34*	*1337*	*2.5*
Cell migration	*34*	*1210*	*2.8*
Positive regulation of metabolic process	32	3111	1
Cellular component organization	30	5578	0.5
Response to organic substance	30	2990	1
Positive regulation of cellular metabolic process	30	2903	1
Cell proliferation	*30*	*1866*	*1.6*
Cell surface receptor signaling pathway	29	2392	1.2
Tissue development	*29*	*1729*	*1.7*
Gene expression	28	5132	0.5
Biosynthetic process	26	5708	0.5
Protein metabolic process	26	5511	0.5
Regulation of cellular component movement	*26*	*852*	*3.1*
Regulation of locomotion	*26*	*841*	*3.1*
Regulation of cell motility	*26*	*773*	*3.4*
Regulation of cell migration	*26*	*730*	*3.6*
Cellular biosynthetic process	25	5533	0.5
Regulation of gene expression	25	4226	0.6
Regulation of signaling	25	3046	0.8
Regulation of cell communication	25	3023	0.8
Regulation of signal transduction	25	2743	0.9

Regulation of cell proliferation	*25*	*1541*	*1.6*
Organic cyclic compound metabolic process	24	5679	0.4
Cellular aromatic compound metabolic process	24	5509	0.4
Heterocycle metabolic process	24	5460	0.4
Nucleobase-containing compound metabolic process	24	5335	0.4
Cell adhesion	*24*	*1208*	*2*
Cellular macromolecule biosynthetic process	22	4672	0.5
RNA metabolic process	22	4331	0.5
RNA biosynthetic process	22	3526	0.6
Transcription, DNA-templated	22	3489	0.6
Regulation of molecular function	22	3060	0.7
Immune system process	22	2343	0.9
Positive regulation of vasculature development	**22**	**175**	**12.6**
Positive regulation of angiogenesis	**22**	**157**	**14**
Cellular protein metabolic process	21	4730	0.4
Regulation of cell differentiation	21	1660	1.3
Regulation of RNA metabolic process	20	3611	0.6
Regulation of RNA biosynthetic process	20	3445	0.6
Regulation of nucleic acid-templated transcription	20	3438	0.6
Regulation of transcription, DNA-templated	20	3421	0.6
Regulation of protein metabolic process	20	2630	0.8
Cellular response to organic substance	20	2175	0.9
Epithelium development	*20*	*1057*	*1.9*

Table 2.8 Top GO biological process annotations associated with the 61 "angiogenesis" genes with larger fold change than Vegfa in OIR + Vehicle vs. RA + Vehicle

GO biological processes associated with the 61 "angiogenesis" genes with larger fold change than Vegfa in OIR + Vehicle vs. RA + Vehicle comparison are identified and those processes with at least 20 genes are displayed in the table. "Selected gene frequency" represents the number of selected "angiogenesis" genes associated with each process. "Total frequency"

represents the total number of mouse genes associated with each process. “Percent of total” is percentage of selected gene frequency out of total frequency. Processes with percentage less than 1.5% are shaded gray, between 1.5% to 10% are italicized and more than 10% are bolded.

KEGG Pathway	Number of genes in pathway	Subset of 61 “angiogenesis” genes with larger fold change than Vegfa that are found in pathway
VEGF signaling pathway	58	None
Calcium signaling pathway	183	Ednra; Tbxa2r
PI3K-Akt signaling pathway	356	Fgf2; Angpt2; Epha2; Pgf; Itgb3; Syk; Itga5; Thbs2; Col4a1; Col4a2; Fn1
MAPK signaling pathway	294	Fgf2; Angpt2; Epha2; Pgf
Focal adhesion	199	Pgf; Itgb3; Itga5; Thbs2; Col4a1; Col4a2; Fn1
Arachidonic acid metabolism	89	None

Table 2.9 Subset of 61 “angiogenesis” genes with larger fold change than Vegfa that are found in pathways down-stream of KEGG “VEGF signaling pathway”
None of the 61 angiogenesis genes with larger fold change than Vegfa in OIR + Vehicle vs. RA + Vehicle comparison are part of the KEGG “VEGF signaling pathway”. The number of selected “angiogenesis” genes that are part of the five pathways downstream of “VEGF signaling pathway” are limited.

KEGG pathway	P-value	Adjusted P-value	General function
mmu04620 Toll-like receptor signaling pathway	5.48E-05	0.001645	Immune
mmu04621 NOD-like receptor signaling pathway	5.48E-05	0.001645	Immune
mmu04622 RIG-I-like receptor signaling pathway	0.001454	0.016353	Immune
mmu04623 Cytosolic DNA-sensing pathway	0.001454	0.016353	Immune
mmu04660 T cell receptor signaling pathway	0.001635	0.016353	Immune
mmu04662 B cell receptor signaling pathway	0.001635	0.016353	Immune
mmu04625 C-type lectin receptor signaling pathway	0.003258	0.024432	Immune
mmu04657 IL-17 signaling pathway	0.002905	0.024432	Immune
mmu04010 MAPK signaling pathway	0.003874	0.025825	Proliferation, migration
mmu04392 Hippo signaling pathway - multiple species	0.005866	0.035198	Proliferation, apoptosis
mmu04014 Ras signaling pathway	0.010398	0.045432	Proliferation, migration
mmu04210 Apoptosis	0.010398	0.045432	Apoptosis
mmu04380 Osteoclast differentiation	0.010601	0.045432	Immune
mmu04668 TNF signaling pathway	0.009687	0.045432	Immune

Table 2.10 KEGG pathways enriched in top 50 TF motifs involved in hypoxia
TFs in the top 50 TF motif groups enriched with OIR + Vehicle vs. RA + Vehicle samples are used to test if any KEGG pathways that the TFs are associated with are observed more than by random chance. 14 KEGG pathways are enriched in this analysis, with eight that are also enriched by KEGG pathway enrichment of hypoxia-modulated genes using *goseq* and *GAGE* (in bold). The enriched pathways function in immune system, cell proliferation, migration, and death.

Drug	Adjusted P-value	Characteristics
Selumetinib	5.38E-12	MAPK kinase inhibitor, cancer drug
Radicicol	2.15E-10	HSP90 inhibitor, cancer drug
Dacomitinib	2.17E-10	EGFR inhibitor, cancer drug
NVP-AUY922	6.00E-10	HSP90 inhibitor, cancer drug
Ruxolitinib	7.63E-10	JAK inhibitor, cancer drug
Geldanamycin	1.06E-09	HSP90 inhibitor, cancer drug
Sunitinib	4.84E-09	Receptor tyrosine kinase inhibitor, cancer drug
BMS-345541	7.07E-09	Anti-inflammatory compound
BMS-387032	8.51E-09	CDK2 inhibitor, cell cycle inhibitor
QL-XII-47	1.22E-08	BTK inhibitor, cancer drug
PHA-767491	1.43E-08	Cdc7/CDK9 inhibitor, cell cycle inhibitor
NVP-TAE684	1.46E-08	NPM-ALK inhibitor, cancer drug
Vorinostat	1.65E-08	HDAC inhibitor, cancer drug
PD-184352	1.65E-08	MAPKK inhibitor, cancer drug
Dovitinib	2.28E-08	FGFR3 inhibitor, cancer drug
TWS-119	2.60E-08	WNT pathway activator
Canertinib	2.73E-08	Tyrosine kinase inhibitor, cancer drug
Pelitinib	2.85E-08	EGFR inhibitor, cancer drug
PF-477736	2.85E-08	CHK1 inhibitor, cancer drug
MK-1775	3.02E-08	Wee1 inhibitor, cell cycle inhibitor

Table 2.11 Top 20 LINCS L1000 studied drugs that show a transcriptome profile *in vitro* that correlates with that of ALG-1001 *in vivo* in OIR mouse

15/20 are cancer drugs with 8 of them being tyrosine kinase inhibitors (in bold). The remaining 5 drugs include three cell cycle inhibitors, an anti-inflammatory compound and a WNT pathway activator.

Drug	Adjusted P-value	Characteristics	Lower cholesterol[1]	Target CNS[2]	Anti-inflammatory[3]
Amiodarone	0.000554	Treat heart rhythm problems			
Simvastatin	0.000923	Statin drug, reduce cholesterol production	X		
Fluphenazine	0.001072	Anti-psychotic drug, block dopaminergic receptors		X	
Fluoxetine	0.001106	Anti-depressant, serotonin reuptake inhibitor		X	
Fludrocortisone Acetate	0.00152	Glucocorticoid replacement			
Fenofibrate	0.00152	Reduce cholesterol level	X		
Dexamethasone	0.001748	Corticosteroid, anti-inflammatory			X
Droperidol	0.002001	Anti-dopaminergic drug		X	
Cerivastatin	0.002077	Reduce cholesterol level	X		
Beta-Estradiol	0.002077	Estrogen steroid hormone			
Sodium Nitroprusside	0.002077	Reduce blood pressure			
Dexfenfluramine	0.0023	Increase extracellular serotonin		X	
Amoxapine	0.002403	anti-depressant		X	
Lovastatin	0.002403	Reduce cholesterol level	X		
Carvedilol	0.002487	Reduce blood pressure			
Fluvastatin	0.002487	Statin drug, reduce cholesterol production	X		
Rosiglitazone	0.002892	anti-diabetic drug, increase insulin response			
Myrtecaine	0.004	Not available			

Buspirone	0.004	Anti-depressant drug		X	
Indomethacin	0.004	Anti-inflammatory drug			X

Table 2.12 Top 20 DrugMatrix drugs that show a transcriptome profile in rat that correlates with that of ALG-1001 in OIR mouse

Drugs that function in lower cholesterol[1], that target CNS[2], or have anti-inflammatory property[3] are labeled by cross.

2.7 References

1. Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M. & Bhattacharya, S. S. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.* **11**, 273–284 (2010).
2. Pennington, K. L. & DeAngelis, M. M. Epidemiology of age-related macular degeneration (AMD): associations with cardiovascular disease phenotypes and lipid factors. *Eye Vis.* **3**, (2016).
3. NEI. Prevalence of Adult Vision Impairment and Age-Related Eye Diseases in America | National Eye Institute. Available at: https://nei.nih.gov/eyedata/adultvision_usa. (Accessed: 19th May 2019)
4. Friedlander, M. Fibrosis and diseases of the eye. *J. Clin. Invest.* **117**, 576–586 (2007).
5. Tang, J. & Kern, T. S. Inflammation in Diabetic Retinopathy. *Prog. Retin. Eye Res.* **30**, 343–358 (2011).
6. Radeke, M. J., Peterson, K. E., Johnson, L. V. & Anderson, D. H. Disease susceptibility of the human macula: Differential gene transcription in the retinal pigmented epithelium/choroid. *Exp. Eye Res.* **85**, 366–380 (2007).
7. Nguyen, Q. *et al.* Management of neovascular age-related macular degeneration: current state-of-the-art care for optimizing visual outcomes and therapies in development. *Clin. Ophthalmol.* 1001 (2015). doi:10.2147/OPTH.S74959
8. Simó, R., Sundstrom, J. M. & Antonetti, D. A. Ocular Anti-VEGF Therapy for Diabetic Retinopathy: The Role of VEGF in the Pathogenesis of Diabetic Retinopathy. *Diabetes Care* **37**, 893–899 (2014).
9. Parikh, R. *et al.* A Multinational Comparison of Anti-Vascular Endothelial Growth Factor Use: The United States, the United Kingdom, and Asia-Pacific. *Ophthalmol. Retina* **3**, 16–26 (2019).
10. Hussain, R. & Ciulla, C. Addressing the Anti-VEGF Treatment Burden. Available at: https://www.reviewofophthalmology.com/article/addressing-the-antivegf-treatment-burden. (Accessed: 19th May 2019)
11. Ehlken, C. *et al.* Switch of anti-VEGF agents is an option for nonresponders in the treatment of AMD. *Eye* **28**, 538–545 (2014).
12. Singer, M. A., Kermany, D. S., Waters, J., Jansen, M. E. & Tyler, L. Diabetic macular edema: it is more than just VEGF. *F1000Research* **5**, (2016).
13. Mackel, M. J., Park, J. Y., Karageozian, H. L. & Karageozian, V. H. Integrin Receptor Antagonists and Their Methods of Use. (2013).
14. Boyer, D. S. *et al.* Integrin Peptide Therapy: A New Class of Treatment for Vascular Eye Diseases - The First Human Experience in DME. *Invest. Ophthalmol. Vis. Sci.* **53**, 1337–1337 (2012).
15. A Clinical Trial Designed to Evaluate the Safety and Efficacy Study of Luminate® in Inducing PVD (Posterior Vitreous Detachment) in Non-Proliferative Diabetic Retinopathy - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02435862. (Accessed: 11th September 2018)

16. A Safety And Efficacy Study Of ALG-1001 In Human Subjects With Wet Age-Related Macular Degeneration - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT01749891. (Accessed: 11th September 2018)
17. Phase 2 Randomized Clinical Trial of Luminate® as Compared to Avastin® in the Treatment of Diabetic Macular Edema - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02348918. (Accessed: 10th May 2019)
18. A Safety And Efficacy Study Of Alg-1001 In Human Subjects With Symptomatic Focal Vitreomacular Adhesion - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02153476. (Accessed: 11th September 2018)
19. Kaiser, P. *et al.* Integrin Peptide Therapy: The First Wet AMD Experience. *Invest. Ophthalmol. Vis. Sci.* **54**, 2177–2177 (2013).
20. Allegro's Luminate Achieves Primary Endpoint in DME Trial. *OIS* (2016). Available at: https://ois.net/allegros-luminate-achieves-primary-endpoint-in-dme-trial/. (Accessed: 19th May 2019)
21. Scott, A. & Fruttiger, M. Oxygen-induced retinopathy: a model for vascular pathology in the retina. *Eye* **24**, 416–421 (2009).
22. Connor, K. M. *et al.* Quantification of oxygen-induced retinopathy in the mouse: a model of vessel loss, vessel regrowth and pathological angiogenesis. *Nat. Protoc.* **4**, 1565–1573 (2009).
23. Stahl, A. *et al.* The Mouse Retina as an Angiogenesis Model. *Invest. Ophthalmol. Vis. Sci.* **51**, 2813–2826 (2010).
24. Shen, J. *et al.* In Vivo Immunostaining Demonstrates Macrophages Associate with Growing and Regressing Vessels. *Invest. Ophthalmol. Vis. Sci.* **48**, 4335–4341 (2007).
25. Joe, S. G., Yoon, Y. H., Choi, J. A. & Koh, J.-Y. Anti-angiogenic effect of metformin in mouse oxygen-induced retinopathy is mediated by reducing levels of the vascular endothelial growth factor receptor Flk-1. *PloS One* **10**, e0119708 (2015).
26. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
27. Kim, C. B., D'Amore, P. A. & Connor, K. M. Revisiting the mouse model of oxygen-induced retinopathy. *Eye Brain* **8**, 67–79 (2016).
28. Campochiaro, P. A. Ocular Neovascularization. *J. Mol. Med. Berl. Ger.* **91**, 311–321 (2013).
29. Vinores, S. A. *et al.* Implication of the hypoxia response element of the vegf promoter in mouse models of retinal and choroidal neovascularization, but not retinal vascular development. *J. Cell. Physiol.* **206**, 749–758 (2006).
30. Vinores, S. A., Xiao, W.-H., Shen, J. & Campochiaro, P. A. TNF-α is critical for ischemia-induced leukostasis, but not retinal neovascularization nor VEGF-induced leakage. *J. Neuroimmunol.* **182**, 73–79 (2007).
31. Guaiquil, V. *et al.* ADAM9 Is Involved in Pathological Retinal Neovascularization. *Mol. Cell. Biol.* **29**, 2694–2703 (2009).

32. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (Accessed: 22nd May 2019)
33. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
34. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
35. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
36. Jiang, H. & Zhan, T. Unit-Free and Robust Detection of Differential Expression from RNA-Seq Data. *Stat. Biosci.* **9**, 178–199 (2017).
37. Zhou, P., Luo, Y., Liu, X., Fan, L. & Lu, Y. Down-regulation and CpG island hypermethylation of CRYAA in age-related nuclear cataract. *FASEB J.* **26**, 4897–4902 (2012).
38. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
39. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
40. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
41. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
42. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
43. Tenenbaum, D. KEGGREST: Client-side REST access to KEGG. *R package version 1.20.0* Available at: http://bioconductor.org/packages/KEGGREST/. (Accessed: 26th September 2018)
44. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).
45. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161 (2009).
46. Shao, Z. ERSSA: Empirical RNA-seq Sample Size Analysis. *Bioconductor* Available at: http://bioconductor.org/packages/ERSSA.
47. Balwierz, P. J. *et al.* ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.* (2014). doi:10.1101/gr.169508.113
48. GeneCards - Human Genes | Gene Database | Gene Search. Available at: https://www.genecards.org/. (Accessed: 16th October 2018)

49. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
50. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
51. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
52. Skelly, D. A. *et al.* Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Rep.* **22**, 600–610 (2018).
53. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
54. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437-1452.e17 (2017).
55. Ganter, B. *et al.* Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* **119**, 219–244 (2005).
56. Ishikawa, K. *et al.* Microarray Analysis of Gene Expression in Fibrovascular Membranes Excised from Patients with Proliferative Diabetic Retinopathy. *Invest. Ophthalmol. Vis. Sci.* (2015). doi:10.1167/iovs.14-15589
57. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).
58. Ishikawa, K. *et al.* Gene Expression Profile of Hyperoxic and Hypoxic Retinas in a Mouse Model of Oxygen-Induced Retinopathy. *Invest. Ophthalmol. Vis. Sci.* **51**, 4307–4319 (2010).
59. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinforma. Oxf. Engl.* **23**, 1846–1847 (2007).
60. Mirabelli, P. *et al.* Genome-wide expression differences in anti-Vegf and dexamethasone treatment of inflammatory angiogenesis in the rat cornea. *Sci. Rep.* **7**, (2017).
61. Newman, A. M. *et al.* Systems-level analysis of age-related macular degeneration reveals global biomarkers and phenotype-specific functional networks. *Genome Med.* **4**, 16 (2012).
62. Whitmore, S. S. *et al.* Altered gene expression in dry age-related macular degeneration suggests early loss of choroidal endothelial cells. *Mol. Vis.* **19**, 2274–2297 (2013).
63. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
64. CaltechDATA. *Caltech Library* Available at: https://www.library.caltech.edu/caltechdata. (Accessed: 9th December 2018)
65. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* **25**, 402–408 (2001).

66. Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* **40**, D1144–D1149 (2012).
67. Potente, M., Gerhardt, H. & Carmeliet, P. Basic and Therapeutic Aspects of Angiogenesis. *Cell* **146**, 873–887 (2011).
68. Chen, L. *et al.* Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget* **9**, 7204–7218 (2018).
69. Krock, B. L., Skuli, N. & Simon, M. C. Hypoxia-Induced Angiogenesis Good and Evil. *Genes Cancer* **2**, 1117–1133 (2011).
70. Tas, S. W., Maracle, C. X., Balogh, E. & Szekanecz, Z. Targeting of proangiogenic signalling pathways in chronic inflammation. *Nat. Rev. Rheumatol.* **12**, 111–122 (2016).
71. Vanhaesebroeck, B., Stephens, L. & Hawkins, P. PI3K signalling: the path to discovery and understanding. *Nat. Rev. Mol. Cell Biol.* **13**, 195–203 (2012).
72. Weis, S. M. & Cheresh, D. A. Tumor angiogenesis: molecular pathways and therapeutic targets. *Nat. Med.* **17**, 1359–1370 (2011).
73. Carlos, T. M. & Harlan, J. M. Leukocyte-endothelial adhesion molecules. *Blood* **84**, 2068–2101 (1994).
74. Langer, H. F. & Chavakis, T. Leukocyte – endothelial interactions in inflammation. *J. Cell. Mol. Med.* **13**, 1211–1220 (2009).
75. Ingber, D. E., Prusty, D., Sun, Z., Betensky, H. & Wang, N. Cell shape, cytoskeletal mechanics, and cell cycle control in angiogenesis. *J. Biomech.* **28**, 1471–1484 (1995).
76. Kiraly, O., Gong, G., Olipitz, W., Muthupalani, S. & Engelward, B. P. Inflammation-Induced Cell Proliferation Potentiates DNA Damage-Induced Mutations In Vivo. *PLoS Genet.* **11**, (2015).
77. Turner, M. D., Nedjai, B., Hurst, T. & Pennington, D. J. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1843**, 2563–2582 (2014).
78. Wheaton, W. W. & Chandel, N. S. Hypoxia. 2. Hypoxia regulates cellular metabolism. *Am. J. Physiol. - Cell Physiol.* **300**, C385–C393 (2011).
79. Choquet, D. & Triller, A. The Dynamic Synapse. *Neuron* **80**, 691–703 (2013).
80. Blomhoff, R. & Blomhoff, H. K. Overview of retinoid metabolism and function. *J. Neurobiol.* **66**, 606–630 (2006).
81. Barczyk, M., Carracedo, S. & Gullberg, D. Integrins. *Cell Tissue Res.* **339**, 269–280 (2010).
82. Senger, D. R. & Davis, G. E. Angiogenesis. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
83. Campbell, N. E. *et al.* Extracellular Matrix Proteins and Tumor Angiogenesis. *Journal of Oncology* (2010). doi:10.1155/2010/586905
84. Staley, J. P. & Woolford, J. L. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr. Opin. Cell Biol.* **21**, 109–118 (2009).
85. Starr, C. R., Pitale, P. M. & Gorbatyuk, M. Translational attenuation and retinal degeneration in mice with an active integrated stress response. *Cell Death Dis.* **9**, 484 (2018).

86. Sun, Y. *et al.* Inflammatory signals from photoreceptor modulate pathological retinal angiogenesis via c-Fos. *J. Exp. Med.* **214**, 1753–1767 (2017).
87. Rockwell, S., Dobrucki, I. T., Kim, E. Y., Marrison, S. T. & Vu, V. T. Hypoxia and radiation therapy: Past history, ongoing research, and future promise. *Curr. Mol. Med.* **9**, 442–458 (2009).
88. Gilany, K. & Vafakhah, M. Hypoxia: a Review. *J. Paramed. Sci.* **1**, (2010).
89. Majmundar, A. J., Wong, W. J. & Simon, M. C. Hypoxia inducible factors and the response to hypoxic stress. *Mol. Cell* **40**, 294–309 (2010).
90. Pe'er, J. *et al.* Hypoxia-induced expression of vascular endothelial growth factor by retinal cells is a common factor in neovascularizing ocular diseases. *Lab. Investig. J. Tech. Methods Pathol.* **72**, 638–645 (1995).
91. van Lookeren Campagne, M., LeCouter, J., Yaspan, B. L. & Ye, W. Mechanisms of age-related macular degeneration and therapeutic opportunities. *J. Pathol.* **232**, 151–164 (2014).
92. Asnaghi, L. *et al.* Hypoxia Promotes Uveal Melanoma Invasion through Enhanced Notch and MAPK Activation. *PLOS ONE* **9**, e105372 (2014).
93. Al-Kharashi, A. S. Role of oxidative stress, inflammation, hypoxia and angiogenesis in the development of diabetic retinopathy. *Saudi J. Ophthalmol.* **32**, 318–323 (2018).
94. Robinson, R., Barathi, V. A., Chaurasia, S. S., Wong, T. Y. & Kern, T. S. Update on animal models of diabetic retinopathy: from molecular approaches to mice and higher mammals. *Dis. Model. Mech.* **5**, 444–456 (2012).
95. Blasiak, J. *et al.* Oxidative Stress, Hypoxia, and Autophagy in the Neovascular Processes of Age-Related Macular Degeneration. *BioMed Res. Int.* **2014**, e768026 (2014).
96. Ambati, J. & Fowler, B. J. Mechanisms of age-related macular degeneration. *Neuron* **75**, 26 (2012).
97. Sato, T. *et al.* Comprehensive gene-expression profile in murine oxygen-induced retinopathy. *Br. J. Ophthalmol.* **93**, 96–103 (2009).
98. Rahman, A. & Fazal, F. Hug Tightly and Say Goodbye: Role of Endothelial ICAM-1 in Leukocyte Transmigration. *Antioxid. Redox Signal.* **11**, 823–839 (2009).
99. Makarev, E. *et al.* Pathway activation profiling reveals new insights into Age-related Macular Degeneration and provides avenues for therapeutic interventions. *Aging* **6**, 1064–1075 (2014).
100. Chen, S.-Y. *et al.* Current concepts regarding developmental mechanisms in diabetic retinopathy in Taiwan. *BioMedicine* **6**, (2016).
101. Suryavanshi, S. V. & Kulkarni, Y. A. NF-κβ: A Potential Target in the Management of Vascular Complications of Diabetes. *Front. Pharmacol.* **8**, (2017).
102. Parveen, A. *et al.* Phytochemicals: Target-Based Therapeutic Strategies for Diabetic Retinopathy. *Molecules* **23**, 1519 (2018).
103. Soetikno, B. T. *et al.* Inner retinal oxygen metabolism in the 50/10 oxygen-induced retinopathy model. *Sci. Rep.* **5**, 16752 (2015).
104. Yokosako, K. *et al.* Glycolysis in Patients with Age-Related Macular Degeneration. *Open Ophthalmol. J.* **8**, 39–47 (2014).

105. Barot, M., Gokulgandhi, M. R. & Mitra, A. K. Mitochondrial Dysfunction in Retinal Diseases. *Curr. Eye Res.* **36**, 1069–1077 (2011).
106. Tarr, J. M., Kaul, K., Chopra, M., Kohner, E. M. & Chibber, R. Pathophysiology of Diabetic Retinopathy. *International Scholarly Research Notices* (2013). doi:10.1155/2013/343560
107. Dorrell, M. I., Aguilar, E., Weber, C. & Friedlander, M. Global Gene Expression Analysis of the Developing Postnatal Mouse Retina. *Invest. Ophthalmol. Vis. Sci.* **45**, 1009–1019 (2004).
108. Bassett, E. A. & Wallace, V. A. Cell fate determination in the vertebrate retina. *Trends Neurosci.* **35**, 565–573 (2012).
109. Farhy-Tselnicker, I. & Allen, N. J. Astrocytes, neurons, synapses: a tripartite view on cortical circuit development. *Neural Develop.* **13**, 7 (2018).
110. Lam, J. D. *et al.* Identification of RUNX1 as a Mediator of Aberrant Retinal Angiogenesis. *Diabetes* **66**, 1950–1956 (2017).
111. Costa, C., Incio, J. & Soares, R. Angiogenesis and chronic inflammation: cause or consequence? *Angiogenesis* **10**, 149–166 (2007).
112. Imhof, B. A. & Aurrand-Lions, M. Angiogenesis and inflammation face off. *Nat. Med.* **12**, 171–172 (2006).
113. Fröhlich, S., Boylan, J. & McLoughlin, P. Hypoxia-Induced Inflammation in the Lung. *Am. J. Respir. Cell Mol. Biol.* **48**, 271–279 (2013).
114. Montgomery, S. T., Mall, M. A., Kicic, A. & Stick, S. M. Hypoxia and sterile inflammation in cystic fibrosis airways: mechanisms and potential therapies. *Eur. Respir. J.* **49**, 1600903 (2017).
115. Yang, Z., Zhao, T., Zou, Y., Zhang, J. H. & Feng, H. Hypoxia Induces Autophagic Cell Death through Hypoxia-Inducible Factor 1α in Microglia. *PLOS ONE* **9**, e96509 (2014).
116. Danese, S., Dejana, E. & Fiocchi, C. Immune Regulation by Microvascular Endothelial Cells: Directing Innate and Adaptive Immunity, Coagulation, and Inflammation. *J. Immunol.* **178**, 6017–6022 (2007).
117. Eltzschig, H. K. & Carmeliet, P. Hypoxia and Inflammation. *N. Engl. J. Med.* **364**, 656–665 (2011).
118. Lim, L. S., Mitchell, P., Seddon, J. M., Holz, F. G. & Wong, T. Y. Age-related macular degeneration. *The Lancet* **379**, 1728–1738 (2012).
119. Cheung, N., Mitchell, P. & Wong, T. Y. Diabetic retinopathy. *The Lancet* **376**, 124–136 (2010).
120. Apte, R. S., Chen, D. S. & Ferrara, N. VEGF in Signaling and Disease: Beyond Discovery and Development. *Cell* **176**, 1248–1264 (2019).
121. Luttun, A., Tjwa, M. & Carmeliet, P. Placental growth factor (PlGF) and its receptor Flt-1 (VEGFR-1): novel therapeutic targets for angiogenic disorders. *Ann. N. Y. Acad. Sci.* **979**, 80–93 (2002).
122. Lieu, C., Heymach, J., Overman, M., Tran, H. & Kopetz, S. Beyond VEGF: Inhibition of the Fibroblast Growth Factor Pathway and Antiangiogenesis. *Clin. Cancer Res.* **17**, 6130–6139 (2011).

123. Nguyen, Q. D. *et al.* Placental growth factor and its potential role in diabetic retinopathy and other ocular neovascular diseases. *Acta Ophthalmol. (Copenh.)* **96**, e1–e9 (2018).
124. Akl, M. R. *et al.* Molecular and clinical significance of fibroblast growth factor 2 (FGF2 /bFGF) in malignancies of solid and hematological cancers for personalized therapies. *Oncotarget* **7**, 44735–44762 (2016).
125. Guaiquil, V. H. *et al.* A Murine Model for Retinopathy of Prematurity Identifies Endothelial Cell Proliferation as a Potential Mechanism for Plus Disease. *Invest. Ophthalmol. Vis. Sci.* **54**, 5294–5302 (2013).
126. Dubrac, A. *et al.* NCK-dependent pericyte migration promotes pathological neovascularization in ischemic retinopathy. *Nat. Commun.* **9**, 3463 (2018).
127. Langmann, T. Microglia activation in retinal degeneration. *J. Leukoc. Biol.* **81**, 1345–1351 (2007).
128. Fischer, F., Martin, G. & Agostini, H. T. Activation of retinal microglia rather than microglial cell density correlates with retinal neovascularization in the mouse model of oxygen-induced retinopathy. *J. Neuroinflammation* **8**, 120 (2011).
129. Zhu, Y. *et al.* Identification of different macrophage subpopulations with distinct activities in a mouse model of oxygen-induced retinopathy. *Int. J. Mol. Med.* **40**, 281–292 (2017).
130. Ziemka-Nalecz, M., Jaworska, J. & Zalewska, T. Insights Into the Neuroinflammatory Responses After Neonatal Hypoxia-Ischemia. *J. Neuropathol. Exp. Neurol.* **76**, 644–654 (2017).
131. Kaur, C., Foulds, W. S. & Ling, E.-A. Hypoxia-ischemia and retinal ganglion cell damage. *Clin. Ophthalmol. Auckl. NZ* **2**, 879–889 (2008).
132. Purves, D. *et al.* The Retina. *Neurosci. 2nd Ed.* (2001).
133. Deng, G. *et al.* Therapeutic Effects of a Novel Agonist of Peroxisome Proliferator-Activated Receptor Alpha for the Treatment of Diabetic Retinopathy. *Invest. Ophthalmol. Vis. Sci.* **58**, 5030–5042 (2017).
134. Moran, E. *et al.* Protective and Antioxidant Effects of PPARα in the Ischemic Retina. *Invest. Ophthalmol. Vis. Sci.* **55**, 4568–4576 (2014).
135. Jeon, C.-J., Strettoi, E. & Masland, R. H. The Major Cell Populations of the Mouse Retina. *J. Neurosci.* **18**, 8936–8946 (1998).
136. ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/home. (Accessed: 16th May 2019)
137. Usui, Y. *et al.* Angiogenesis and Eye Disease. *Annu. Rev. Vis. Sci.* **1**, 155–184 (2015).
138. Stewart, M. W. Extended Duration Vascular Endothelial Growth Factor Inhibition in the Eye: Failures, Successes, and Future Possibilities. *Pharmaceutics* **10**, (2018).
139. Paul, M. K. & Mukhopadhyay, A. K. Tyrosine kinase – Role and significance in Cancer. *Int. J. Med. Sci.* **1**, 101–115 (2004).
140. Page, T. H., Smolinska, M., Gillespie, J., Urbaniak, A. M. & Foxwell, B. M. J. Tyrosine kinases and inflammatory signalling. *Curr. Mol. Med.* **9**, 69–85 (2009).

141. Hoppe, G. *et al.* Comparative systems pharmacology of HIF stabilization in the prevention of retinopathy of prematurity. *Proc. Natl. Acad. Sci.* **113**, E2516–E2525 (2016).
142. Yang, X., Dong, X., Jia, C. & Wang, Y. Profiling of genes associated with the murine model of oxygen-induced retinopathy. *Mol. Vis.* **19**, 775–788 (2013).
143. Oubaha, M. *et al.* Senescence-associated secretory phenotype contributes to pathological angiogenesis in retinopathy. *Sci. Transl. Med.* **8**, 362ra144-362ra144 (2016).
144. Hunter, A. *et al.* DNA methylation is associated with altered gene expression in AMD. *Invest. Ophthalmol. Vis. Sci.* **53**, 2089–2105 (2012).
145. Antonetti, D. A. *et al.* Diabetic Retinopathy. *Diabetes* **55**, 2401–2411 (2006).
146. Guerrero, P. A. & McCarty, J. H. TGF-β Activation and Signaling in Angiogenesis. *Physiol. Pathol. Angiogenesis - Signal. Mech. Target. Ther.* (2017). doi:10.5772/66405
147. Darland, D. C. & D'Amore, P. A. Blood vessel maturation: vascular development comes of age. *J. Clin. Invest.* **103**, 157–158 (1999).
148. Cines, D. B. *et al.* Endothelial Cells in Physiology and in the Pathophysiology of Vascular Disorders. *Blood* **91**, 3527–3561 (1998).
149. Retina | anatomy. *Encyclopedia Britannica* Available at: https://www.britannica.com/science/retina. (Accessed: 18th January 2019)
150. Kolb, H. Roles of Amacrine Cells. in *Webvision: The Organization of the Retina and Visual System* (eds. Kolb, H., Fernandez, E. & Nelson, R.) (University of Utah Health Sciences Center, 1995).
151. Oh, S., Shin, S. & Janknecht, R. ETV1, 4 and 5: An Oncogenic Subfamily of ETS Transcription Factors. *Biochim. Biophys. Acta* **1826**, 1–12 (2012).
152. Jia, J. *et al.* AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation. *Microvasc. Res.* **105**, 103–108 (2016).
153. Ota, M. & Sasaki, H. Mammalian Tead proteins regulate cell proliferation and contact inhibition as transcriptional mediators of Hippo signaling. *Dev. Camb. Engl.* **135**, 4059–4069 (2008).
154. Murphy, M., Chatterjee, S. S., Jain, S., Katari, M. & DasGupta, R. TCF7L1 Modulates Colorectal Cancer Growth by Inhibiting Expression of the Tumor-Suppressor Gene EPHB3. *Sci. Rep.* **6**, 28299 (2016).
155. Werth, D. *et al.* Proliferation of human primary vascular smooth muscle cells depends on serum response factor. *Eur. J. Cell Biol.* **89**, 216–224 (2010).
156. Bergeron, R. *et al.* Chronic activation of AMP kinase results in NRF-1 activation and mitochondrial biogenesis. *Am. J. Physiol.-Endocrinol. Metab.* **281**, E1340–E1346 (2001).
157. Takano, Y. *et al.* Study of drug effects of calcium channel blockers on retinal degeneration of rd mouse. *Biochem. Biophys. Res. Commun.* **313**, 1015–1022 (2004).
158. Metoki, T. *et al.* Study of Effects of Antiglaucoma Eye Drops on N-Methyl-d-Aspartate-Induced Retinal Damage. *Jpn. J. Ophthalmol.* **49**, 453–461 (2005).
159. Kandpal, R. P. *et al.* Transcriptome analysis using next generation sequencing reveals molecular signatures of diabetic retinopathy and efficacy of candidate drugs. *Mol. Vis.* **18**, 1123–1146 (2012).

160. Baruch, K. *et al.* PD-1 immune checkpoint blockade reduces pathology and improves memory in mouse models of Alzheimer's disease. *Nat. Med.* **22**, 135–137 (2016).
161. Green, E. M. *et al.* A small-molecule inhibitor of sarcomere contractility suppresses hypertrophic cardiomyopathy in mice. *Science* **351**, 617–621 (2016).
162. Rao, A. *et al.* Inhibition of ileal bile acid uptake protects against nonalcoholic fatty liver disease in high-fat diet–fed mice. *Sci. Transl. Med.* **8**, 357ra122-357ra122 (2016).
163. DNA Sequencing Costs: Data. *National Human Genome Research Institute (NHGRI)* Available at: https://www.genome.gov/27541954/dna-sequencing-costs-data/. (Accessed: 7th February 2019)
164. Cui, Y. & Paules, R. S. Use of transcriptomics in understanding mechanisms of drug-induced toxicity. *Pharmacogenomics* **11**, 573–585 (2010).
165. Sujirakul, T. *et al.* Multimodal imaging of central retinal disease progression in a 2 year mean follow up of Retinitis Pigmentosa. *Am. J. Ophthalmol.* **160**, 786–98.e4 (2015).

2.8 Appendix

Appendix 2.8.1 OIR retina transcriptome at P17 compared to prior literature

In the past, several groups have used high-throughput transcriptome methods (RNA-seq and microarray) to study biology of OIR retina, with limited discovery of modulated genes and biological processes (**Table 2.6**). In contrast to these earlier studies, the present study improved in three key aspects. First, the we measured the transcriptome at height of hypoxia-induced angiogenesis (postnatal day P17)[97], rather than at onset of hypoxia (P12[141], P12.5[58], P13[142], and P14[143]). This approach allow us to measure the full effect of retinal hypoxia and avoid inadvertently characterizing effects of hyperoxia, which is still measurable at P12.5[58] and possibly later. Second, we included measurements of the effect of both hypoxia and a therapeutic agent (Hoppe *et al.*[141] only examined effect of two drugs in OIR model, while Ishikawa *et al.*[58], Yang *et al.*[142], and Oubaha *et al.*[143] only evaluated OIR transcriptome). Finally, we employed a relatively large number of replicates (n=11-12) to discern the transcriptome response due to retinal hypoxia (in contrast to the small number of replicates previously used: 3 in Ishikawa *et al.*[58], 3 in Yang *et al.*[142], and 2 in Oubaha *et al.*[143]). The large sample size allows us to detect significantly more hypoxia-regulated genes with higher confidence (as sample size increases, **Figure S2.8A** shows more genes discovered, **Figure S2.8B-C** show higher confidence in the discovery) and enables us to perform in-depth analyses using GO biological process, KEGG pathway and TF enrichments that are either not performed before[58,143] or are of limited capability[142] in previous studies.

Qualitatively, prior studies of OIR retina at P12.5[58] and P13[142] found consistent evidence of elevated angiogenesis and inflammatory transcript levels. In Ishikawa *et al.* (moderately high transcriptome correlation, **Figure 2.17A**)[58], a significant number of 53 elevated genes identified at P12.5 are associated with vasculogenesis and inflammation. Similarly, at P13, Yang *et al.*[142] discovered 62 elevated genes that enriched 13 GO biological processes, including "angiogenesis" and "immune response". Intriguing, all of these processes are represented in our study except "glycolysis" and "glucose metabolic process", which are known to be activated during hypoxia response[145]. Further analysis suggest activation of glycolysis may be transient and has already receded at P17: out of the six glycolysis genes found to be elevated at P13, only three are still elevated at P17 and with much smaller fold changes compared to at P13.

Among the prior transcriptome studies of OIR retina, none of them reported a suppression of neural genes and only Oubaha *et al.*[143] reported elevated levels of apoptosis and inflammation genes at P14, which are also observed at P17 in our dataset ("apoptosis" and inflammation KEGG pathways such as "NF-kB signaling", **Figure 2.6**). The lack of evidence for activation of cell death at the early time points (P12.5, P13) and their presence at the later time points (P14, P17) suggest transcriptome signature of cell death may not be detectable until hypoxia-activated physiological responses are fully established post-P13.

Comparative analysis also suggest VEGF-mediated blood vessel growth may be in the decline at P17. First, "VEGF signaling pathway" was not enriched with elevated genes at P17 (adjusted p-value=0.609, 15/50 pathway genes elevated). Next, we looked at a time

series (P12 to P21) qPCR study[97] that analyzed gene levels of 94 genes (including Vegfa) between OIR and control retina (Pearson's r=0.91 with exclusion of outlier Nrp2 gene, **Figure 2.18B**). In the time course study, Vegfa level gradually increased and reached peak level at P16 before falling 48.9% on P17, followed by further decrease on subsequent days. Intriguingly, Tgfb1, a blood vessel maturation[146] (and possible angiogenesis inhibitor[147]) gene, was also found to be co-activated in OIR mice and reached peak level one day later on P17 (42.7% increase from P16) before a gradual decline. Collectively, these evidence suggest the molecular activation of blood vessel formation is likely already in decline at P17 with the tissue transitioning to post-angiogenesis vessel maturation[148].

Clearly, hypoxia-response in the OIR model is dynamic with transient elevation of glycolysis, angiogenesis and inflammation genes (supported by Sato *et al.*) and delayed signs of apoptosis and neural stress. While the Sato *el al.* study was able to capture some of these transcriptome dynamics by focusing on 94 genes, future studies should explore multiple time points in a single high-throughput study to fully capture the transient changes in the retinal hypoxia response.

Appendix 2.8.2 Validation of tissue composition analysis

To validate retinal cell marker-based tissue composition analysis (**Figure 2.19, 2.20**), we employ two well cited scRNA-seq studies of mouse brain (cortex and hippocampus)[51] and mouse heart and immune cells[52]. First, we analyzed markers derived from the brain, as both the brain and retina are formed from the same neural tissue and presumably share similar cell

types[149]. From the brain study, we obtained markers associated with nine major brain cell types (3/330 or <1% overlap with retinal cell markers), which showed endothelial cell and microglia cell numbers are likely expanded in the OIR mice (**Figure 2.21**, pericyte is not annotated in the brain study). Additionally, >80% of interneuron (related to amacrine cells in the retina[150]) markers showed negative fold change in the OIR mice, providing additional evidence that selected neuronal cells are depleted in the OIR retina.

Similarly, with the heart and immune cell markers (5/96 or 5.2% overlap with retinal cell markers, none are of the same cell type), we saw >80% of endothelial cell, macrophages (related to microglia in the retina) and pericytes markers have positive fold change in OIR mice (**Figure 2.22**). Not surprisingly, there wasn't any cell types with a large number of negative fold change markers, as no neuronal cells was annotated in this data set. Altogether, based on mostly orthogonal gene sets derived from non-retinal tissues, these results show consistent evidence of vessel and immune-related cell expansion and neuronal cell depletion in the OIR retina.

Appendix 2.8.3 Suppressed genes in DR retina compare to OIR retina

In this study, we found suppressed genes in hypoxia response are over-represented with ones associated with neural system processes, indicative of a detectable disruption in normal neural function and development. Similarly, in human retinal diseases that occur much later in life, there are also clear signs of neural damage from hypoxia, neovascularization and inflammation[119]. Supporting these observations, we performed a GO biological process

enrichment of suppressed genes in human DR tissue[56] and found neural system processes such as “ion transport” and “nervous system development” are enriched (8/36 GO biological processes enriched overlap with those enriched with suppressed genes in hypoxia response, p-value=3.00e-05). However, the DR study also has major caveat – it compared non-neural fibrovascular membrane tissue collected from DR patients to neural retina collected from normal donors. Since non-neural fibrovascular membrane presumptively contains fewer cells of neural origin, this could also explain the lower level of neural mRNA in the DR samples.

Appendix 2.8.4 Correlation between OIR and human disease retina among subset of hypoxia-modulated genes

We set out to find if any biological process-specific transcriptome features of hypoxia response in OIR mice are correlated with the human retinal diseases. In particular, we individually analyzed fold change correlation of genes among “immune system process”, “angiogenesis”, “cell death”, “metabolic process”, and “nervous system development” GO biological processes, all of which are enriched in GO biological process enrichment except “metabolic process” (included to encompass 20 other enriched metabolism-related processes).

In this analysis (**Table 2.5**), we found when limited to “immune system process” and “cell death” genes, correlations generally improved across the different disease cohorts, demonstrating OIR mouse may be a good model for these disease processes. Next, correlation with “nervous system development” genes improved within the AMD cohorts, but slightly dropped with DR (still relatively high at r=0.2999). This observation suggest the

modulated neural system in OIR retina are able to capture some signatures of the neural retina disruption in AMD[118] and DR[119]. With "angiogenesis" genes, correlation reaches the highest level in our analysis to r=0.4073 between OIR mice and DR, but dropped for most of the AMD cohorts, including those that develop angiogenesis. Here, the relatively high correlation with DR shows the OIR model is useful to model angiogenesis in these patients, but the low correlation to AMD cohorts, especially those that develop neovascularization, is surprising (perhaps due to differences in angiogenesis between the two diseases, **Figure 2.23**). As a control, we also performed correlation analyses with non-statistically significant genes in hypoxia response, which produced near zero correlations across all disease cohorts and served as a validation of this analysis approach.

Appendix 2.8.5 ALG-1001 affects endothelial cell migration and adhesion to ECM

ALG-1001's effect on the transcriptome of OIR retina is largely in agreement with *in vitro* HUVEC cell culture studies (**Chapter 3**) and is further supported by endothelial cell adhesion and migration studies (Campochiaro group, JHU). First, in HUVECs treated with ALG-1001 under hypoxia condition, a statistically significant number of angiogenesis genes were selectively down-regulated. Additional analysis of ALG-1001-modulated genes *in vitro* found strong evidence of selective down-regulation of genes that encode proteins that function in the ECM and in modulating cell adhesion. Of these, we found fibronectin, two collagens, two laminin, and four integrin subunit genes are down regulated after drug treatment.

Transcriptome suppression of cell adhesion and migration genes manifested in detectable disruption of these cellular functions in *in vitro* endothelial cell studies. In these experiments, our colleagues at the Campochiaro group found ALG-1001 specifically inhibited endothelial cell migration on fibronectin-coated and less strongly on vitronectin-coated surfaces (personal communication). Additionally, ALG-1001 treatment weakly inhibited endothelial adhesion to both fibronectin-coated and vitronectin-coated surfaces.

Appendix 2.8.6 Comparison of ALG-1001 transcriptome profile to anti-VEGF treatment

In our search, we found nearly all of the transcriptome studies of anti-VEGF drugs are investigation of Bevacizumab in context of cancer treatments, which are not directly comparable to ocular diseases due to increased complexity of cancer heterogeneity and pathophysiology. There is one recent study[60] that tested anti-VEGF antibody and a control anti-IgG antibody in a rat model that develops trauma-induced angiogenesis and inflammation in the cornea. While analyzing the rat cornea dataset, we were surprised to see that after 2 days of treatment, no gene level is changed by anti-VEGF antibody compare to anti-IgG antibody (this observation was also reported by the original authors). Interestingly, the authors also noted that the effect of anti-VEGF antibody is quite modest in these models, where drug treatment was found to reduce neovascularization by less than 15-20%.

Next, we compared the fold changes measured in the rat cornea study with the present study, limited to the genes modulated by ALG-1001 in the OIR mice. In this analysis, we found almost no correlation between the two datasets (Pearson's r=-0.0324), but interestingly, 74%

of genes that have negative fold change after ALG-1001 treatment also have negative fold change after anti-VEGF treatment in rat cornea, indicating some level of consistency in the drug-induced transcriptome change (**Figure S2.11**). On the other hand, whereas ALG-1001 mainly suppressed transcript level, there are non-statistically significant evidence that suggest anti-VEGF treatment appears to elevate gene level more than gene suppression (1298 probes "up" and 1040 probes "down" with p-value<0.05 by anti-VEGF antibody, compared to 566 genes "up" and 1036 genes "down" with p-value<0.05 by ALG-1001 in OIR mice).

Appendix 2.8.7 TF groups responsible for effect of ALG-1001

ISMARA was used to predict TFs that may be involved in ALG-1001 associated transcriptome changes in OIR mice. In this analysis, no TF was enriched with Z-score>1.0, which is typically used as the statistical significance cutoff (**Table 2.3**). While weakly supported by evidence, we set out to evaluate the top 50 TF groups identified with the understanding that the results are not statistically sound. In the top 50 TF groups, we identified at least 12 TF groups with known function related to the cell proliferation, angiogenesis, hypoxia-response, and inflammatory processes (labeled in **Figure 2.15**). Out of the 12 identified, 9 have suppressed target gene level, consistent with *edgeR* that shows majority of drug-induced transcriptome changes are suppressed. In addition, 8/12 of these TF motifs are ranked very high on the enrichment list, representing more than half of the top 15 TF groups enriched.

We also compared the top 50 TF motifs identified in both hypoxia and ALG-1001-modulated transcriptomes and find 11 TF groups that are on both lists (bold TFs in **Tables 2.2 and 2.3**). For 8 of these 11 TF groups, the direction of target gene level is reversed by ALG-1001 treatment and will be the focus of the following discussion. Notably, the 8 TF groups include Fos[86], Jun[86] and several members of the ETS protein family[151] (Erg, Etv1 and Etv5) that are well known regulator of cell proliferation and angiogenesis (Fos and Jun are down-stream of VEGF signaling[152]). In addition, the other TFs also play important function in various aspects of the hypoxia response: Tead1 (regulates cell proliferation and migration[153]), Tcf7l1 (regulates cell cycle progression[154]), Srf (stimulates cell proliferation and differentiation[155]), and Nrf1 (regulates metabolism and cell growth[156]). Altogether, TF analysis provides additional evidence that ALG-1001 treatment in OIR mice moderated some of the hypoxia-initiated transcriptional changes.

Appendix 2.8.8 Limitations of the L1000 study

There are two important caveats with the L1000 study: a) drug exposure were done in human cancer cell lines, and b) inference of expression in >90% of genes. First, the LINCS L1000 transcriptome study was performed with human cancer cell lines, instead of with animal model or human primary cells. In our studies, there was a large gap in drug-induced transcriptome change between animal and cell culture studies (for example, 93.8% of ALG-1001-modulated genes in OIR mice were suppressed, but only 48.8% in HUVEC and 53.4% in PBMC, **Chapter 3**). Thus, the results between the present *in vivo* study and the *in vitro* cancer cell study may not be directly comparable. A second major limitation of the LINCS

L1000 study is it measured the expression of 978 genes and then used their expression profile to infer the expression of 9196 additional genes, all of which were used to generate the list of compound-regulated genes. Since gene expression was not directly measured for a majority of the genes, analysis performed with the LINCS L1000 dataset should be interpreted with caution.

Appendix 2.8.9 Literature survey shows transcriptome studies tend to use diseased model, but not control animals for drug research

A small-scale survey of the recent literature on transcriptome study of ocular drug treatment shows our study of combined control and diseased animals is quite unconventional. Of the four studies that we reviewed, all of them probed the drug effect in diseased models, but not in control animals[141,157–159]. Similarly, a broader survey outside the ocular field arrived at the same observation[160–162]. While limited in scope, this survey shows current transcriptome studies of therapeutics tend to mainly focus on identifying the drug effect in diseased models. In these studies, high cost of high-throughput approaches may be one explanation for the lack of research in control animals. However, as sequencing cost continues to drop[163], it becomes desirable to study the drug-induced transcriptome change not only in diseased model to identify therapeutic effects, but also in control animals to assess the drug's safety profile[164].

Appendix 2.8.10 Comparison of ALG-1001 transcriptome profile with DrugMatrix compounds

Compared to the L1000 dataset, the DrugMatrix study is significantly smaller in scale, yet it measured the effect of drugs in *in vivo* rat animals. Among the drugs included in the DrugMatrix dataset, 89 compounds shared transcriptome profile with ALG-1001 with statistical significance. Of those, we focused on characterizing the top 20 hits (**Table 2.12**) through a manual search of their function. Unlike the L1000 study, there wasn't an abundance of anti-cancer drugs in the top compounds. Instead, there are five cholesterol lowering drug, six central nervous system (CNS)-targeting drugs, two anti-inflammatory drugs and others of miscellaneous functions. Importantly, most of the enriched drugs' mechanism of action remains unknown (several developed before era of molecular biology), reducing their utility in uncovering ALG-1001's mechanism of action.

Appendix 2.8.11 Enrichment of TFs involved in hypoxia response

When *ISMARA* was applied to study OIR + Vehicle vs. RA + Vehicle, 104 motif groups are enriched with high confidence (defined as ones with Z-score>1.0, top 50 shown in **Table 2.2**). To make interpretation feasible, analysis was limited to the top 50 enriched TF groups. Through a manual review of their biological function, we found many of them are involvement in cell proliferation, angiogenesis, inflammation, metabolism, hypoxia-response and neuronal development (labeled in **Figure 2.11**, activity profile shown in **Figure 2.12-13**). For an objective analysis, we employed the KEGG pathway database to identify pathways that contain one or more TFs in the top 50 enriched TF groups. Of the 156 KEGG pathways that contain at least one TF tested by *ISMARA*, we found 60 pathways contained at least one TF in the top 50 enriched TF groups. Next, we tested whether any of these 60

KEGG pathways are over-represented in the top 50 TF groups than one would expect by random chance and identified 14 KEGG pathways with statistical significance (**Table 2.10**). Of the 14 enriched-KEGG pathways, 8 (**Table 2.10**, bolded) are also enriched by *goseq* or *GAGE* with elevated genes in OIR mice, which serves as a validation of this line of TF functional analysis. Additionally, all 14 TF-enriched KEGG pathways are strikingly relevant to the immune system, apoptosis, and cell proliferation and migration processes, demonstrating the active involvement of these pathways and their associated TFs in regulating retinal hypoxia response.

While many of the top TF groups are relevant to hypoxia response, it is worth noting that over half of them (27/50, groups not labeled in **Figure 2.11** and shown in **Table 2.2**) are either *not* well established to regulate a putative hypoxia process or are *uncharacterized* altogether. In the top 20 TF groups alone, there are 4 groups where we did not find a connection with hypoxia: regulators of estrogen response (Esrra and Esrrb), iron homeostasis (Spic) and two mostly uncharacterized putative regulators: Nhlh1 and Hsfy2. These observations suggest the retinal transcriptome regulation in response to low oxygen is complex, with involvement of TFs of seemingly unrelated function or are uncharacterized altogether.

Appendix 2.8.12 Limitations of OIR mouse model

The extensive literature on OIR (over 15,000 publications since 1994) documents a robust animal model for DR[23]. While the OIR model has been useful in studying the development

of abnormal angiogenesis in the retina, it is important to bear in mind the differences between the mouse model and human diseases. First, in the mouse model, hypoxia and angiogenesis are transient events in a young animal, while the human retinal diseases are chronic and typically occur much later in life[95]. Second, in contrast to the rapidly activated and quickly resolved angiogenesis in the mouse OIR model[23], human retinal diseases and their associated physiological changes develop over a long period of time and typically do not get better over time[118,165]. Furthermore, the pathophysiology of the human retinal diseases is highly complex and includes involvement of many contributing factors related to both environmental stress and genetics predisposition[6,118]. Thus, while the OIR model may capture hypoxia-activated retinal angiogenesis and other processes to a certain extent, the human diseases are driven by additional factors that are not part of the mouse model. As a result, it is important to view our findings in the context of the limited ability of the mouse OIR model to capture the characteristics of human retinal diseases.

Appendix 2.8.13 Fold change of hypoxia-modulated genes after ALG-1001 treatment

To further investigate ALG-1001's hypoxia-modulating effect, we broadened the analysis to all 6208 hypoxia-modulated genes at P17 and analyzed their corresponding fold change after ALG-1001 treatment. Intriguingly, we found a moderate and negative correlation between the two comparisons (r=-0.258, **Figure S2.12**). In particular, of the 3139 genes elevated, over 67% (2117/3139) have negative fold-change after ALG-1001 injection (r=-0.471). Further investigation of the 2117 genes using GO biological process enrichment showed many of the angiogenesis and inflammatory processes (**Figure S2.13A**) previously enriched with

hypoxia-elevated genes in OIR mice, are recovered (877 out of 1157, **Figure S2.13C**). To confirm ALG-1001 specifically suppressed genes associated with hypoxia-elevated processes, we repeated for 1000 times a randomly selection of 2117 genes from the 3139 hypoxia-elevated genes and calculated the number of processes that are enriched with the selected genes and that are also previously enriched with the 3139 genes. In this analysis, we see the 877 common processes observed in **Figure S2.13C** is highly unlikely by random chance, as only 1/1000 selections showed a stronger overlap (across all 1000 simulations: mean=661, standard deviation=72 overlapping processes). Additionally, a control enrichment with the remaining 1022 genes (elevated in OIR mice and with positive fold change with ALG-1001 treatment) only identified 77 of the previous 1157 enriched biological processes (GO biological processes in **Figure S2.13B**, overlap in **Figure S2.13D**).

2.9 Appendix Figures

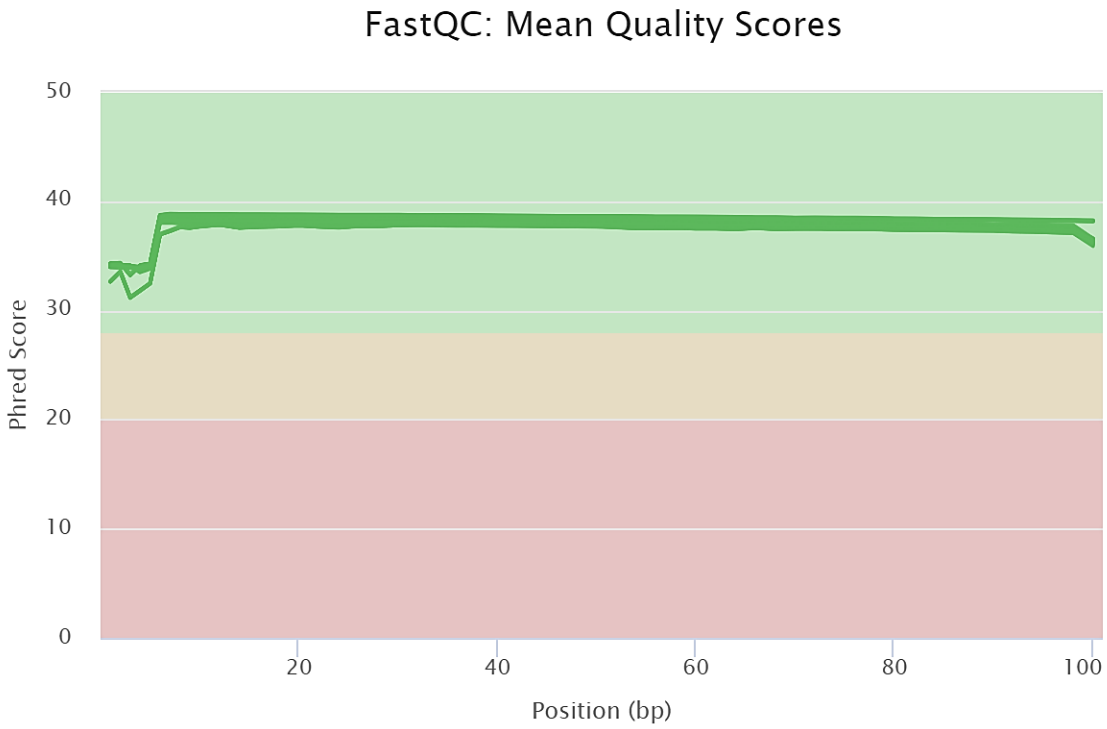

Figure S2.1 *FastQC* mean sequencing quality score
The averaged sequencing quality (Phred score, generated by *FastQC*[32]) at each base in the reads generated for each sample. Each green line detonates a sample and the majority of them overlap for the majority of the entire 100 base pair region. Phred score shows good sequencing quality across all samples.

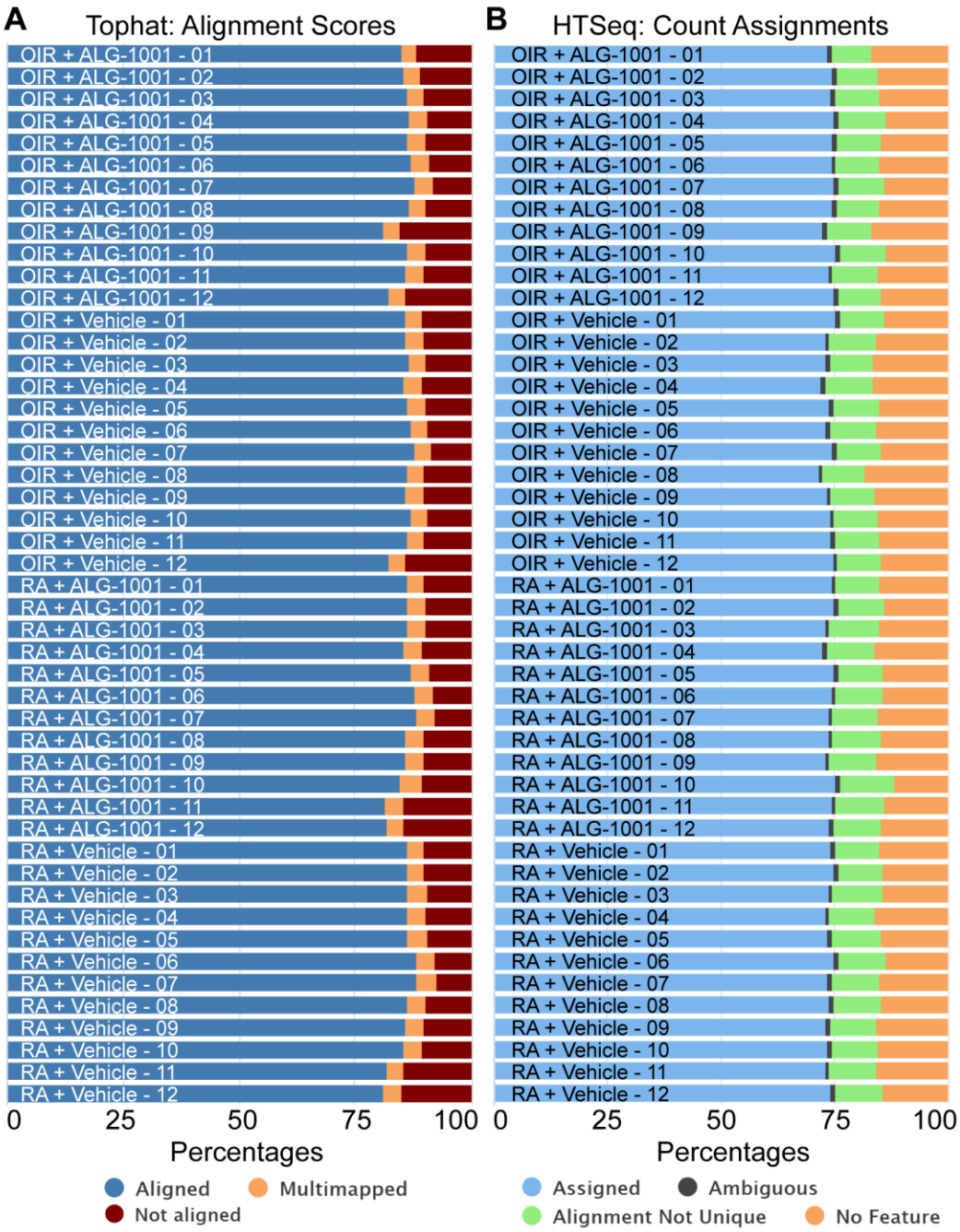

Figure S2.2 *Tophat2* and *HTSeq-count* statistics
Label in bar indicates sample condition and sample ID, 12 samples per condition. **A.** At least 85% of reads are mapped by *Tophat2* in each of 48 samples. **B.** *HTSeq-count* uniquely assigned (blue bar) between 72-76% of a sample's aligned reads to a gene.

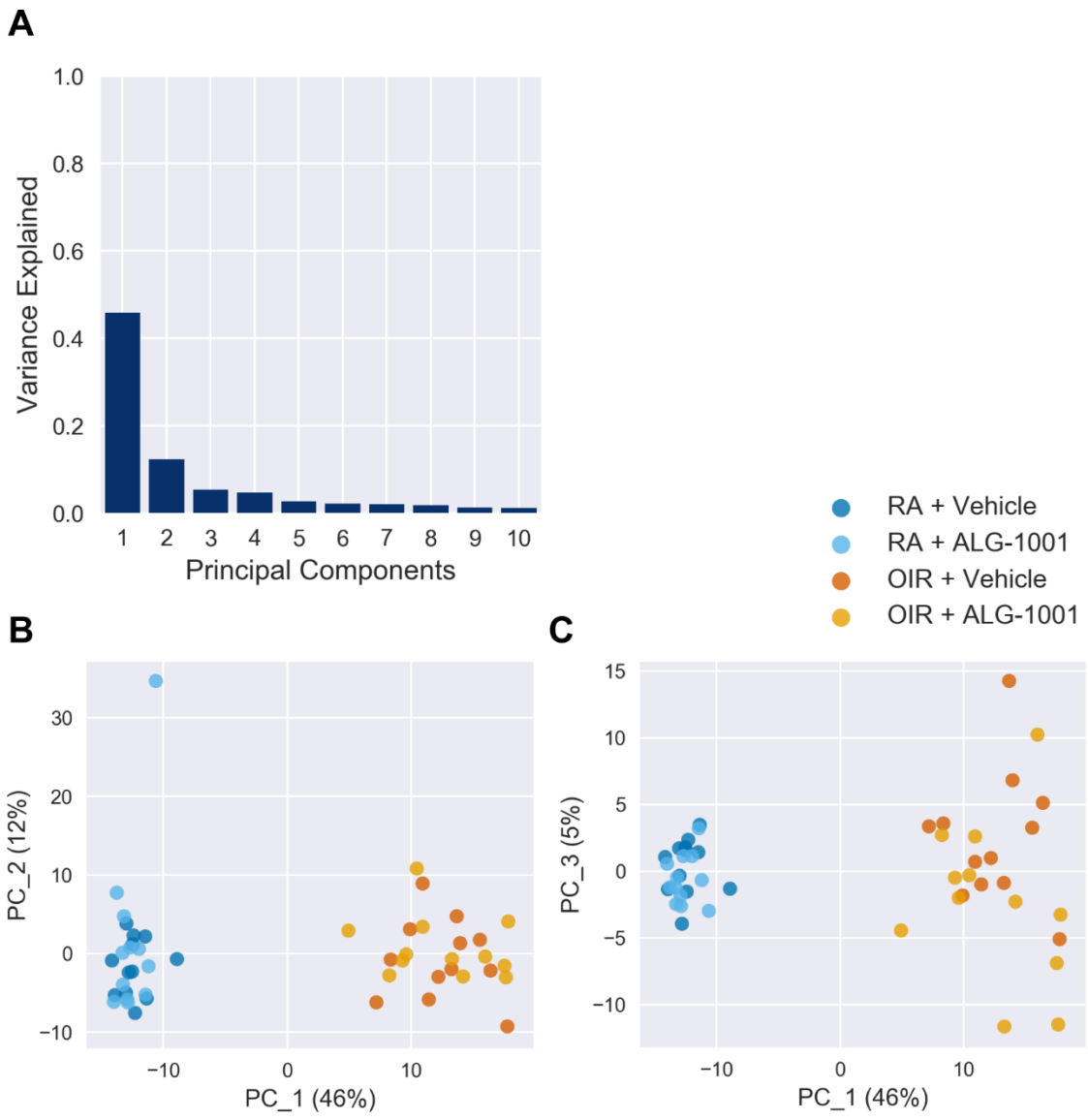

Figure S2.3 PCA of all samples reveals an outlier
PCA_{All} with all 48 retinal samples. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC1 and PC2, PC1 and PC3, respectively. PC_# is principal component number and percent of variance explained is shown in axis label. $PC1_{All}$ strongly separates OIR from RA samples. One RA + ALG-1001 sample appears to be an outlier on $PC2_{All}$.

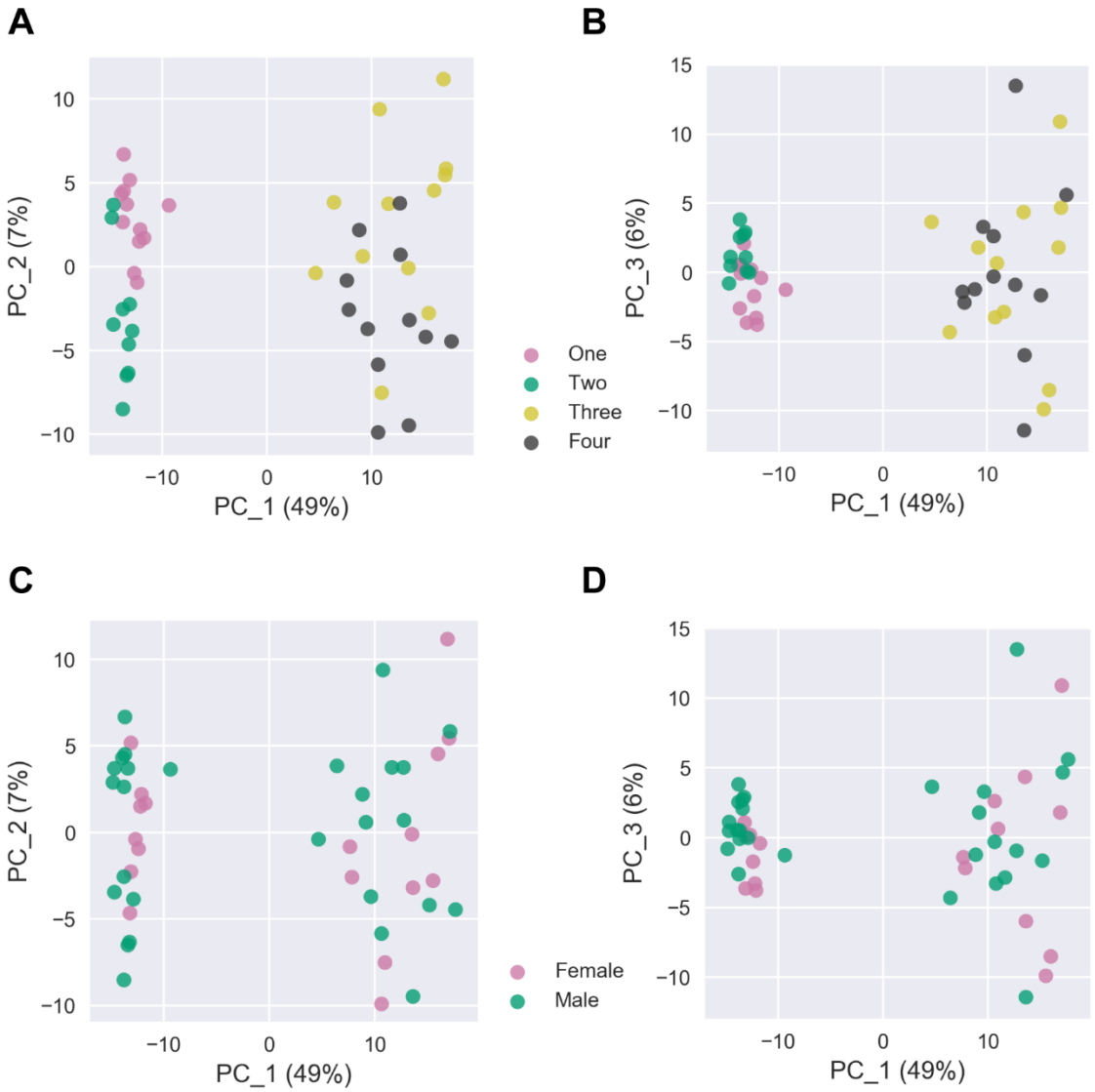

Figure S2.4 PCA with outlier pair removed and color indicating litter or gender
$PCA_{No\text{-}outlier}$ with all samples except the outlier and its fellow eye sample. PC_# is principal component number. Percent of variance explained are shown in axis label. Color-code for sample-litter association (**A-B**) or sample-gender association (**C-D**). (**A, C)** Samples in PC1 and PC2. (**B, D**) Samples in PC1 and PC3. $PCA_{No\text{-}outlier}$ shows weak separation based on litter, indicating presence of litter-specific effect, which is blocked out in *edgeR* analysis. No clear separation associated with gender after exclusion of gender-specific genes (see **Methods**).

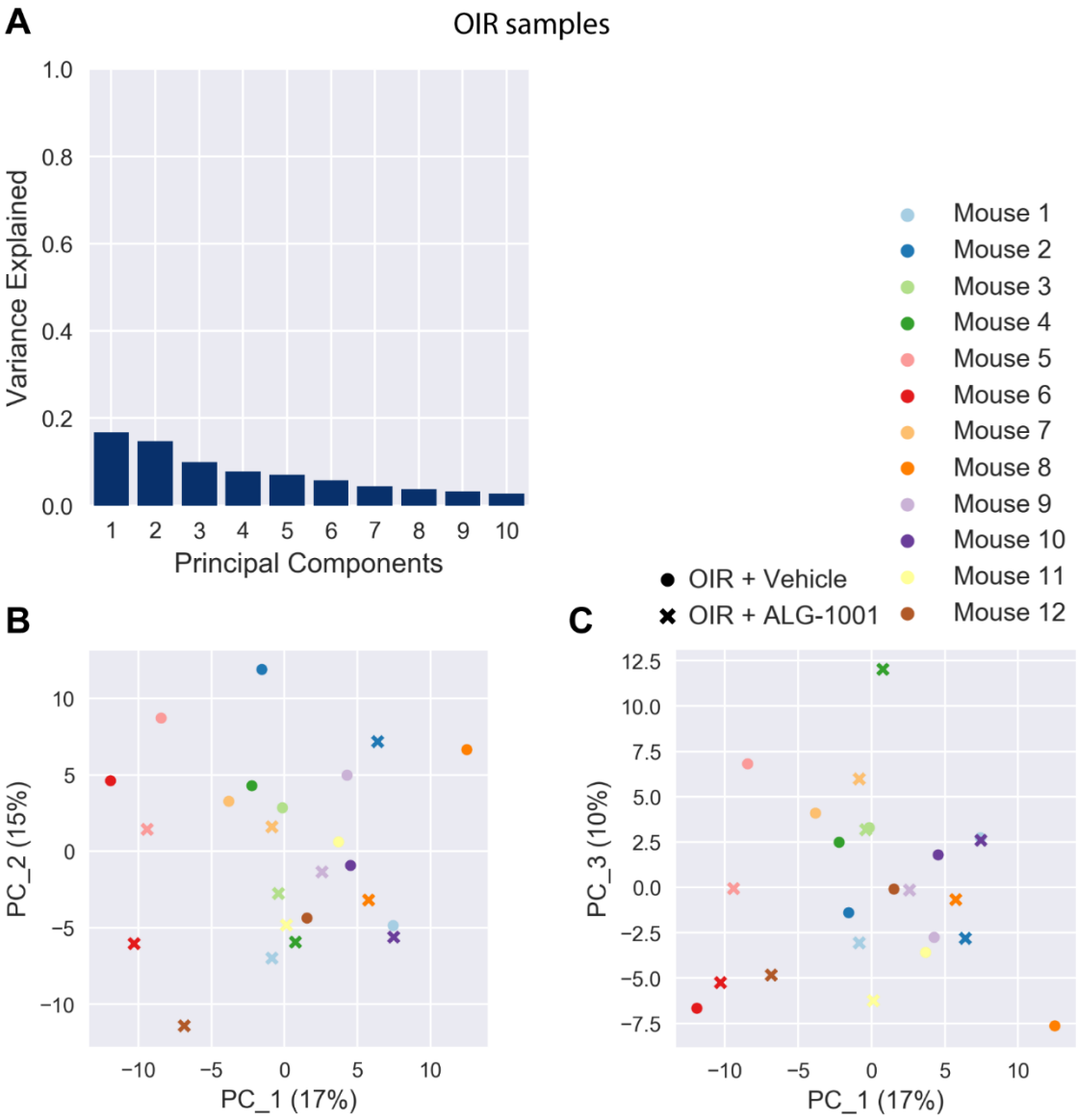

Figure S2.5 PCA with OIR samples
PCA_{OIR} with 24 OIR samples. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC1 and PC2, PC1 and PC3, respectively. PC_# is principal component # and percent of variance explained is shown in axis label. Symbols indicate condition while colors indicate animal.

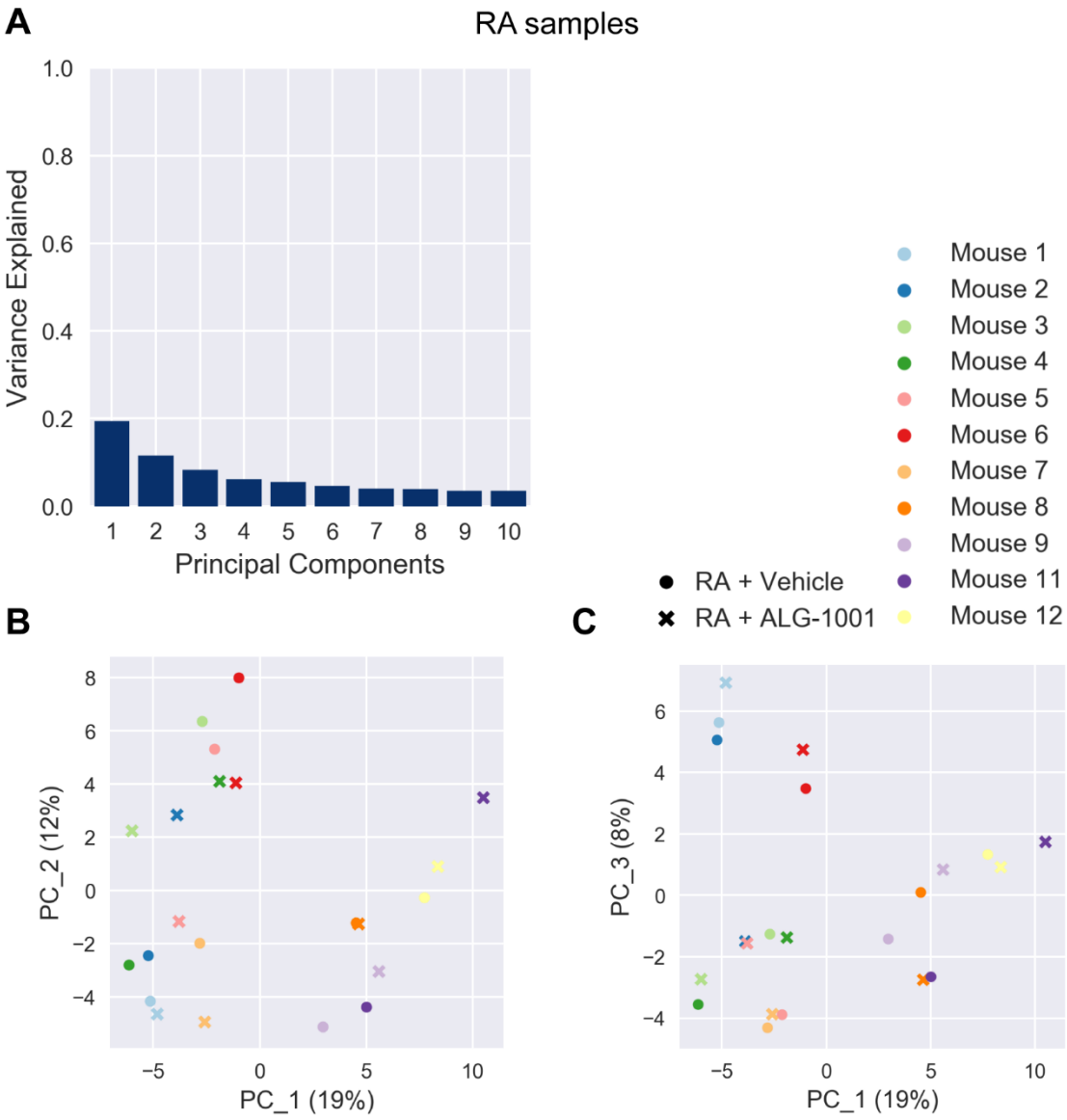

Figure S2.6 PCA with RA samples
PCA_{RA} with 22 RA samples excluding outlier pair (Mouse 10). **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC1 and PC2, PC1 and PC3, respectively. PC_# is principal component # and percent of variance explained is shown in axis label. Symbols indicate condition while colors indicate animal.

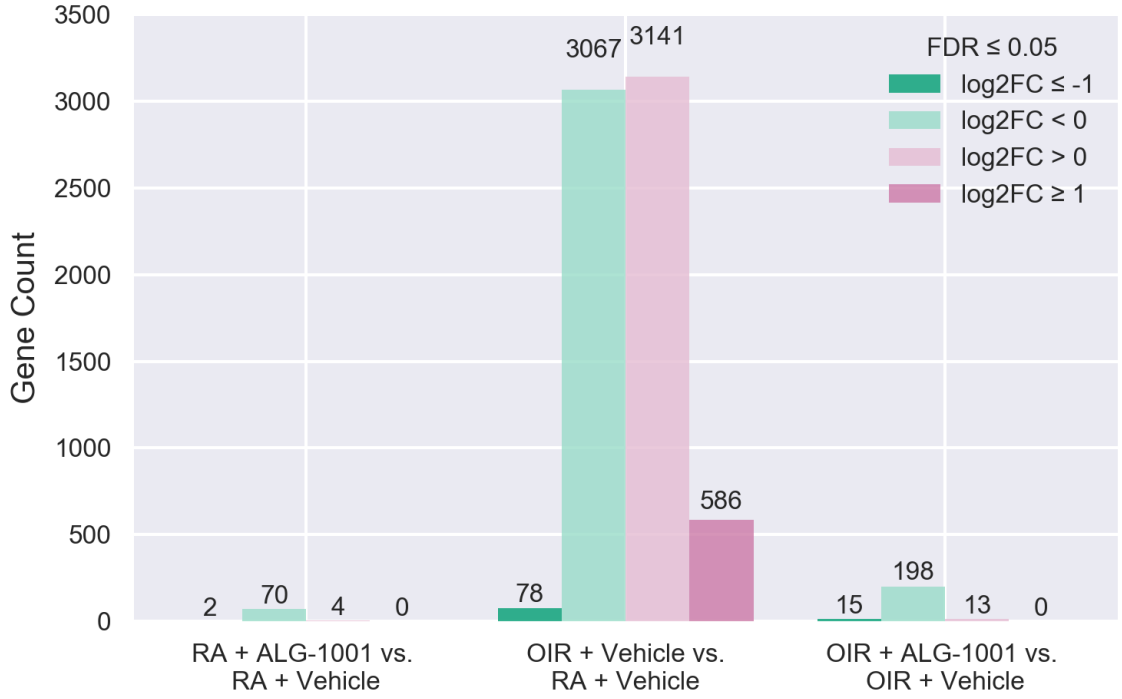

Figure S2.7 Number of modulated genes in each *edgeR* comparison
The number of genes with statistically significant change (false discovery rate (FDR)<0.05) in each of the three *edgeR* comparisons are plotted as bar graph. The number of genes in each bar is shown above the bar. Log2FC is log2-fold-change.

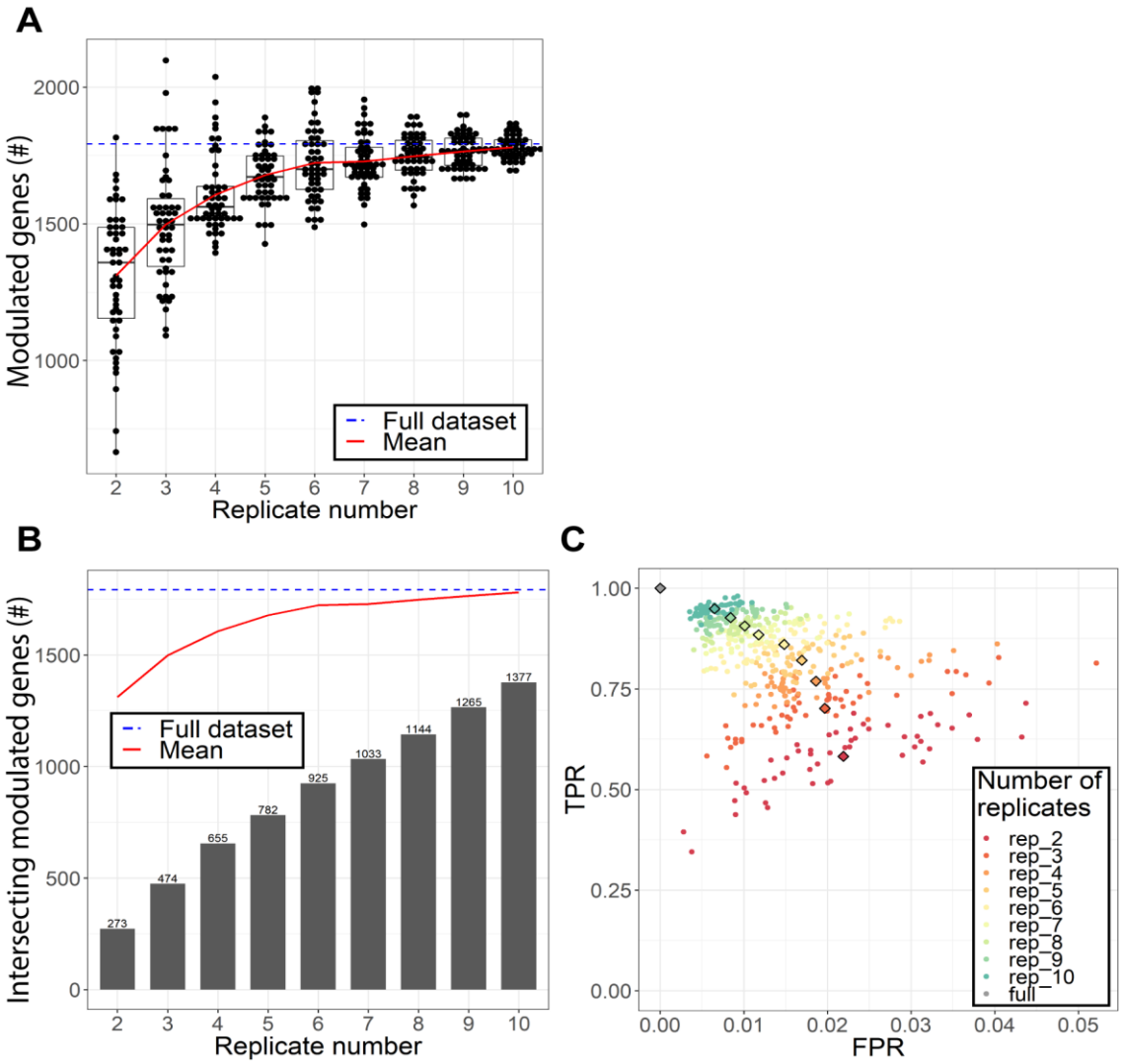

Figure S2.8 Sample size analysis with ERSSA

ERSSA is used to determine if sufficient sample size has been used to maximize discovery of genes with statistically significant change in the comparison of OIR + Vehicle vs. RA + Vehicle. **A.** Plot shows the measured number of hypoxia-modulated gene at each replicate number with 50 combinations per sample size and |log2-fold-change| cut off=0.5. Subsampling results show improved discovery of modulated genes as sample size increases up to the full dataset. **B.** Plot shows number of modulated genes found in all subsamples at each replicate number. As sample size increases, the consensus improves among the subsamples, indicating more reliable discovery. **C.** Plot shows individual and averaged false positive rate (FPR) and true positive rate (TPR) among the subsamples with full dataset discovery as the ground truth. TPR tend to increase, while FPR tend to decrease as sample size increases.

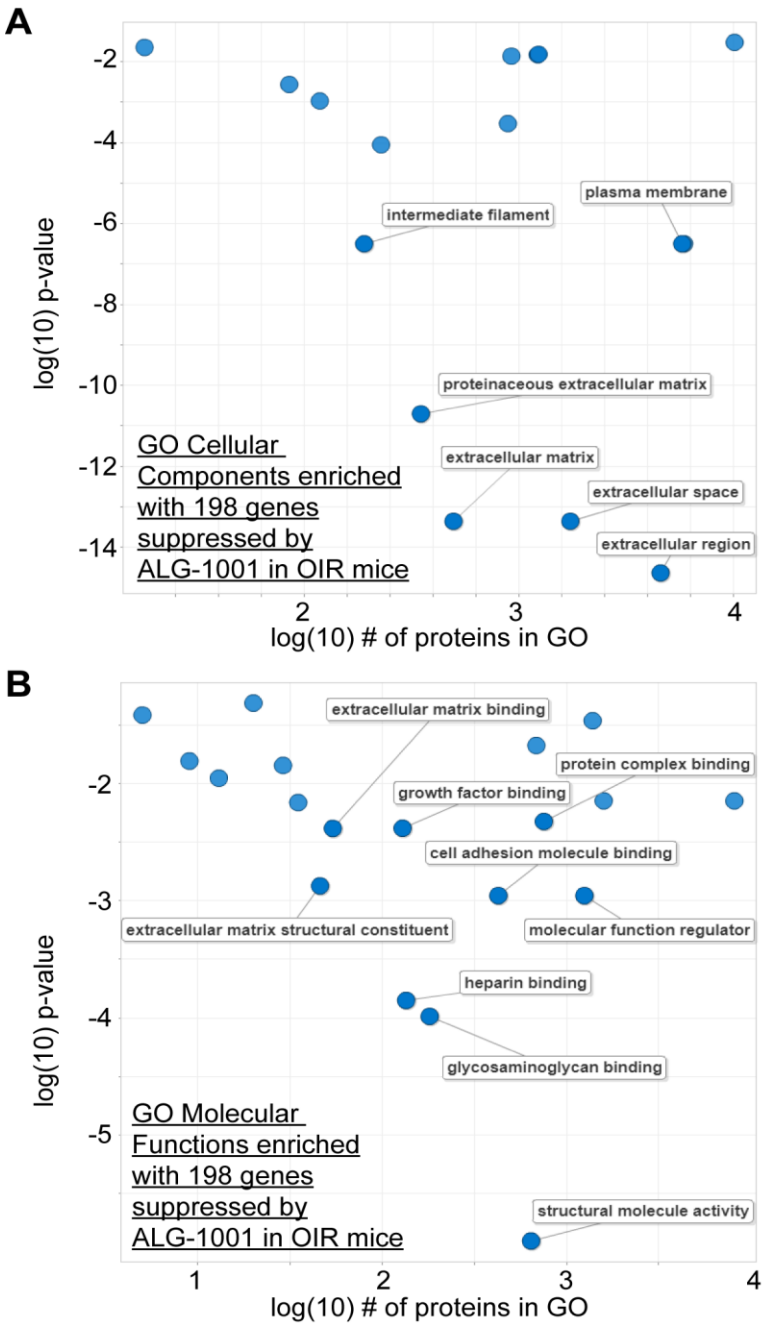

Figure S2.9 Cellular components and molecular functions enriched with 198 genes suppressed by ALG-1001 in OIR mice

Suppressed genes after ALG-1001 treatment OIR mice relative to vehicle treated control are submitted for GO cellular component (**A**) and molecular function (**B**) enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Selected biologically relevant processes are labeled.

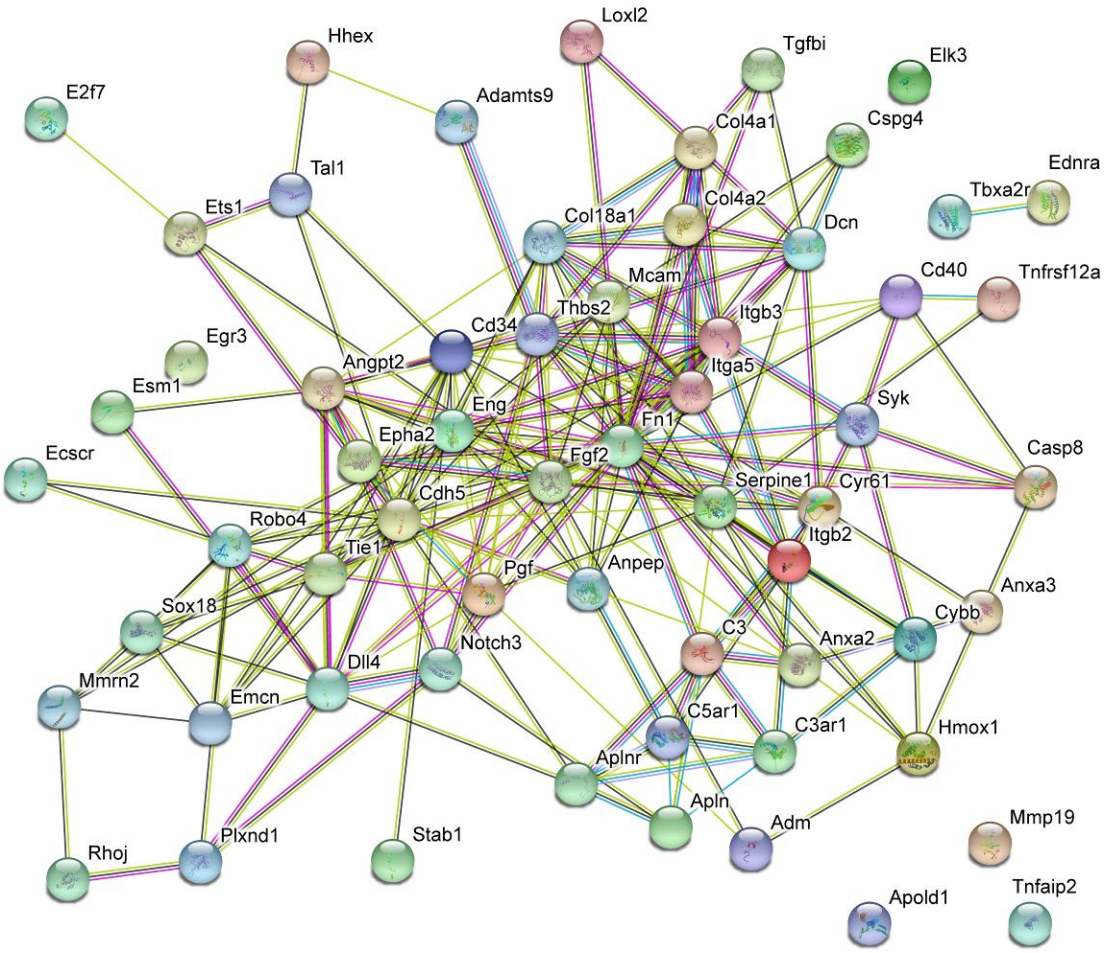

Figure S2.10 *STRING* connection map of 61 selected "angiogenesis" genes
Selected "angiogenesis" genes have larger fold change than Vegfa in OIR mice (OIR + Vehicle vs. RA + Vehicle). The 61 genes are submitted to *STRING* to visualize connection between gene-coded proteins. *STRING* connection map shows the 61 "angiogenesis" genes are closely connected.

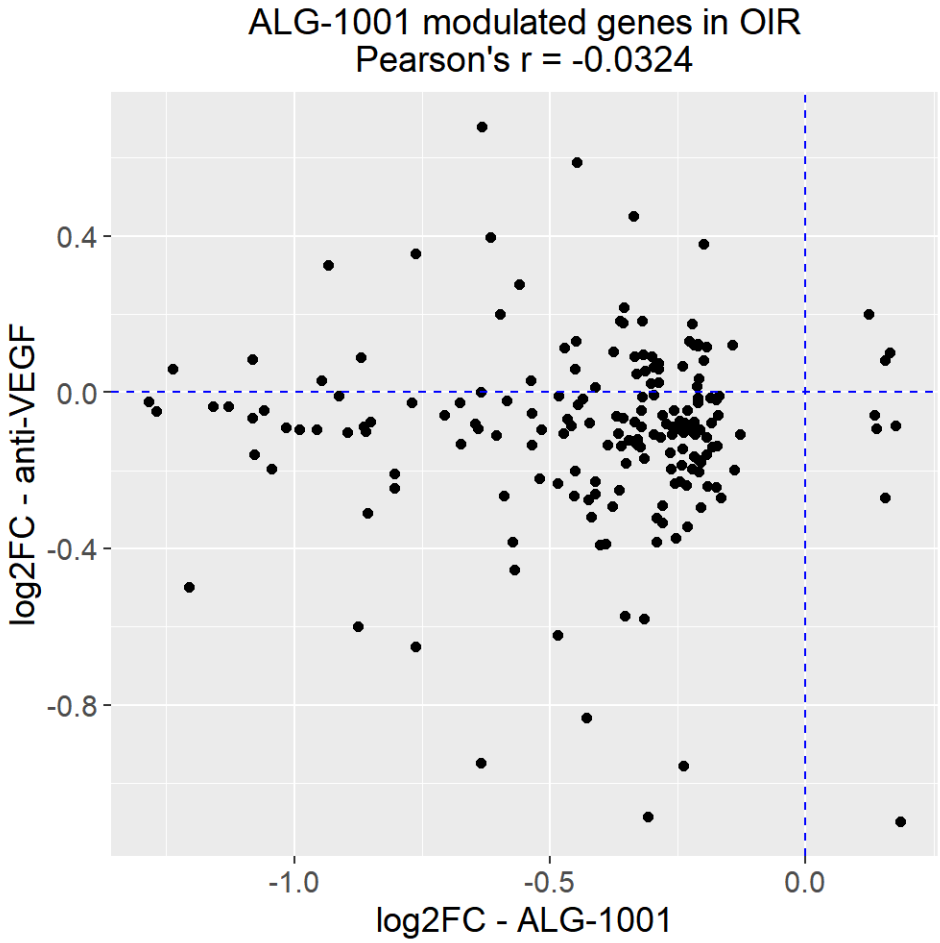

Figure S2.11 Fold change comparison with anti-VEGF drug in rat cornea
Log2-fold-change of modulated genes after ALG-1001 treatment in OIR mice relative to vehicle treated fellow control are plotted on the X-axis. The same gene's log2-fold-change after anti-VEGF drug treatment in rat cornea trauma study is plotted on the Y-axis. Dashed blue lines indicate Y=0 and X=0. Pearson's correlation coefficient is indicated in the plot title. While little to no correlation is detected between the datasets, 74% of genes suppressed by ALG-1001 also have negative fold change in anti-VEGF treated animals.

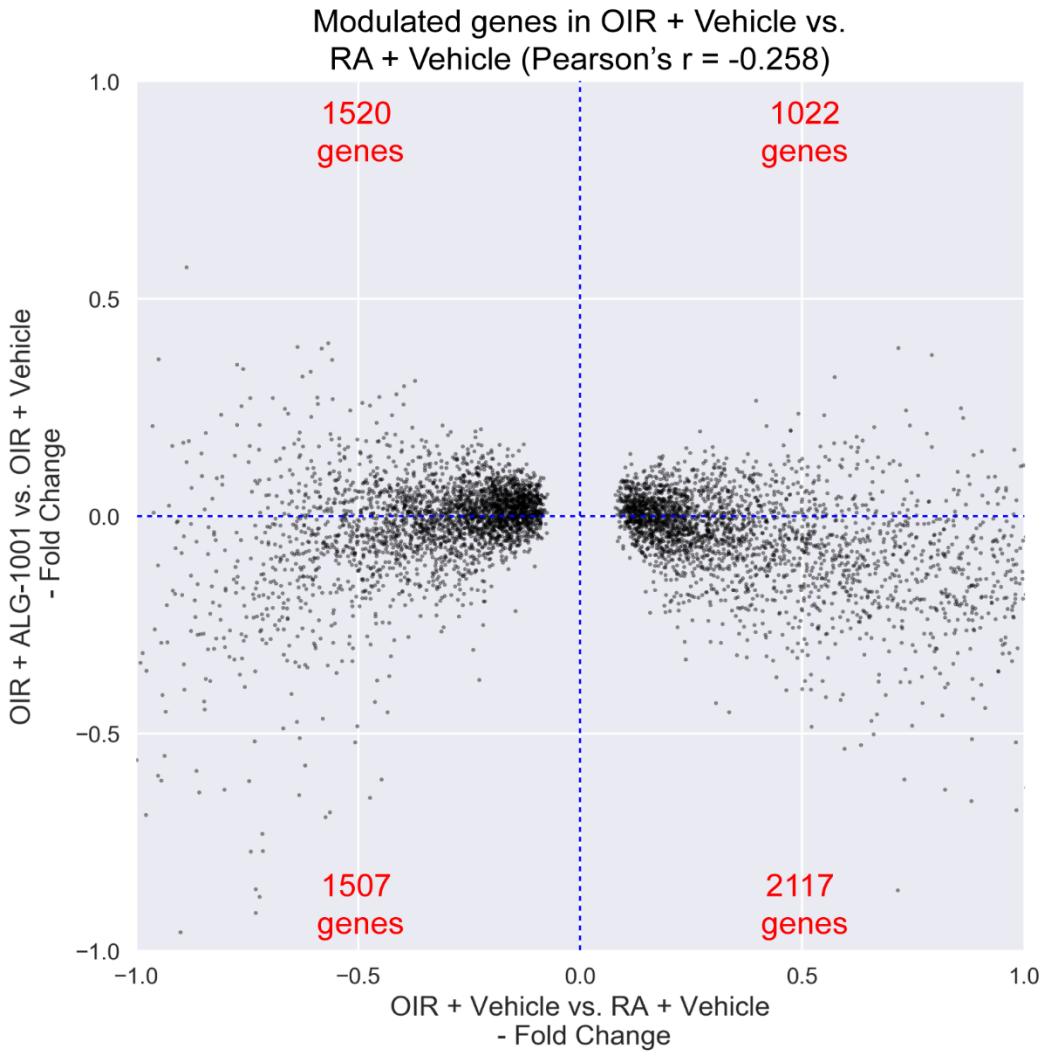

Figure S2.12 Fold change of hypoxia-response genes after ALG-1001 treatment
Each dot represents a gene modulated in OIR + Vehicle vs. RA + Vehicle comparison, with the gene's log2-fold-change plotted on the x-axis. The same gene's fold change after ALG-1001 treatment in OIR mice relative to vehicle control is plotted on the y-axis. Plot window is limited to the (-1, 1) range on both axes. Pearson's r value is measured as -0.258. The plot can be divided into four quadrants, as indicated by the blue dashed lines with 1520 genes in the top left, 1507 in the bottom left, 1022 in the top right, and 2117 in the bottom right quadrant. Note the numbers do not add up to the number of modulated genes in OIR + Vehicle vs. RA + Vehicle, because some genes failed to pass TPM cutoff in OIR + ALG-1001 vs. OIR + Vehicle comparison.

Figure S2.13 GO biological process enrichment of elevated genes in OIR mice
Among hypoxia-elevated genes (OIR + Vehicle vs. RA + Vehicle), 2177 genes with negative fold change (**A**) and 1022 genes with positive fold change (**B**) after ALG-1001 treatment (OIR + ALG-1001 vs. OIR + Vehicle) are submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Selected biologically relevant processes are labeled. **C-D.** Overlap of enriched processes in (**A** or **B**, respectively) with those enriched with all 3141 hypoxia-elevated genes.

Chapter 3. Transcriptome response to ALG-1001 treatment in human endothelial and immune cell cultures

3.1 Introduction

Degenerative diseases of the retina affect millions of Americans, who suffer from vision loss and even blindness[1–5]. The current standard of care uses antibodies to disrupt the Vascular Endothelial Growth Factor (VEGF) pathway to suppress the disease-worsening retinal neovascularization[5]. However, the current treatment has many shortcomings, raising the need to develop a second line of effective therapeutic (see **Chapter 1.1 and 2.1** for more detail)[6,7]. To address these unmet needs, ALG-1001 was developed by Allegro Ophthalmics, LLC. and is currently undergoing clinical testing with promising results[8]. In multiple phase I and II clinical trials, patients who received intraocular injections of the drug experienced favorable vision improvement with no drug-related adverse effect to date[9,10]. Despite its success in the clinical trials, the precise mechanism of action of ALG-1001 remains unclear.

In this study, we employed two human primary cell culture models to study the transcriptome regulation driven by ALG-1001 treatment. First, human umbilical vein endothelial cells (HUVECs) are used as a model for vascular endothelial cells, which play a primary role during neovascularization through their active proliferation, migration and tube formation[11]. Retinal neovascularization has long been associated with retinal disease progression[12,13] and previous *in vivo* mouse studies have showed ALG-1001 has anti-angiogenic property (see

Chapter 1.2 and 2.1 for more detail). Second, human peripheral blood mononuclear cells (PBMCs) consisting of mostly lymphocytes (T cells, B cells and NK cells) and monotypes are used as a model for immune cells typically found in sustaining a pro-inflammatory state in the retinal diseases (see **Chapter 1.1** for detail)[14–16].

For each cell culture model, we treated the cells with ALG-1001 and then captured the drug-related transcriptome response using RNA-sequencing (RNA-seq)[17]. In addition, we exposed HUVECs to 3% O_2 hypoxia to induce a disease-relevant growth environment and probed the gene expression with and without ALG-1001 co-treatment. By analyzing the expressed transcriptome, we identified the genes, biological processes and pathways regulated by ALG-1001 in both models and by hypoxia in HUVECs. Using the available expression dataset, we also identified transcription factors (TFs) that may be involved in the expression regulation. Results from these analyses suggest ALG-1001 can specifically modulate angiogenesis genes in endothelial cells and inflammatory genes in immune cells, bringing us closer to a broader understanding of the drug's therapeutic effect in retinal diseases.

3.2 Materials and Methods

Cell culture

Pooled HUVECs at passage 2 (P2) were purchased from American Type Culture Collection (ATCC) and grown to P5-P6 for experiments. Cell cultures were kept in a 37˚C humidified incubator with pH controlled by 5% CO_2 and media exchanged every two days. The cells

were grown in Vascular Cell Basal Medium (ATCC) with Endothelial Cell Growth Kit-BBE (ATCC) and Antibiotic Antimycotic Solution (Sigma-Aldrich) added according to the manufacturer's instruction. For hypoxia treatment, the culture plates were moved to a hypoxia chamber (Biospherix) with O_2 concentration set to 3% and placed inside the cell culture incubator. For ALG-1001 treatment, the peptide (CPC Scientific) was dissolved in water at a high concentration and then added to the growth media to the final concentration for cell exposure. 500μM of ALG-1001 was used for RNA-seq experiments while a concentration series of ALG-1001 was used for other assays. Equal volume of vehicle water was also added to control samples. After appropriate amount of hypoxia and drug exposure, the cells were collected and assayed (see **Figure 3.1** for more details).

For PBMC study, freshly collected PBMCs from a human donor were purchased from Zen-Bio. Immediately upon arrival, the cells were washed with growth media (RPMI (Thermo Fisher) with 5% heat-inactivated fetal bovine serum (Thermo Fisher) and Antibiotic Antimycotic Solution) and stained with trypan blue to confirm >90% cell viability. Cells were then allowed to rest in 6-well, Ultra-low attachment plates (Corning) overnight in a 37˚C humidified incubator with pH controlled by 5% CO_2. The next morning, cells were treated for 8 hours with 100μM of ALG-1001 (peptide was first dissolved in water to a high concentration and then added to the growth media to the final treatment concentration). The same volume of vehicle water was added to control samples. At the end of 8 hours treatment period, both suspended and adherent cells were collected for RNA-seq sample preparation.

<u>Dead and live staining</u>

Dead and live staining was performed with HUVECs after treatment with a concentration series of ALG-1001. Cells were cultured in 12-well plates (Falcon) and followed the treatment regime specified in **Figure 3.1**. After treatment, LIVE/DEAD Viability/Cytotoxicity Kit (Invitrogen) was used to simultaneously capture the status of both dead and live cells. At start of the assay, cells were washed once with Dulbecco's Phosphate-Buffered Saline (PBS, Thermo Fisher) before incubation for 30 minutes at 37˚C in the media-dye mixture. At end of incubation, cells were washed with PBS and imaged on a Zeiss Axiovert 25CFL microscope. FIJI software package was used to merge fluorescent signals into one image with red indicating dead cells and green indicating live cells. Phase contrast images were captured on the same microscopy system with bright-field illumination. A 10x Lysis Buffer (Thermo Scientific Pierce) was used as the lethal control and added to the wells to working concentration for 15 minute incubation before staining.

WST-1 assay

WST-1 assay was performed with HUVECs treated with a concentration series of ALG-1001. Cells were cultured in 96-well plates (Falcon) and followed the treatment regime specified in **Figure 3.1**. At start of the assay, WST-1 cell proliferation reagent (Roche) was added to the cell culture media at the recommended working concentration for four hours incubation at 37˚C. At end of incubation, the assay intensity was measured on a Flexstation 3 microplate reader (440nm and 690nm). The baseline absorption from the background and media was subtracted to isolate the contribution of the cells. All test well measurements were normalized as a percentage of the mean control well value. One-way ANOVA tests were

performed between control and treated cells in GraphPad Prism 7.00 with the statistical significance threshold set at p-value < 0.05.

RNA-seq sample preparation

For RNA-seq studies, HUVECs were cultured on 6-well plates (Falcon) and PBMCs on 6-well, Ultra-low attachment plates (Corning). At end of treatment, HUVECs were washed once with PBS without calcium and magnesium (Thermo Fisher) and collected using Accutase (Stemcell Technologies). For PBMC, suspended cells were first collected and then combined with adherent cells isolated with the same technique as in HUVECs. Total RNA was extracted using RNeasy Mini Kit (Qiagen) and contaminating DNA removed using TURBO DNA-free kit (Thermo Fisher). RNA quality was measured using Bioanalyzer (Agilent Genomics); all samples were found to have at least 8.40 RNA integrity number. RNA-seq libraries were then prepared using NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs Inc.) and sequenced on the HiSeq 2500 (Illumina) to generate on average 12.5 million, single-end, 100 base pair reads per sample.

RNA-seq analysis

Quality control

FASTQ files generated by RNA-seq were submitted to *FastQC* for quality control. All samples showed good sequencing quality and were submitted for read alignment. Alignment was performed using *Tophat2* to the *hg38* genome and transcriptome references with *Bowtie 2* sensitivity level set to very sensitive[18]. After read alignment, gene counts were quantified

using *HTSeq-count* using the intersection-strict model[19]. *FastQC, tophat2,* and *HTSeq-count* reports were visualized using *multiQC* software (**Figures 3.2-3.7**)[20]. All samples exhibited good alignment property except replicate 5 of the ALG-1001+hypoxia condition in HUVEC batch 1 study, and as a result this sample was removed from all subsequent analysis.

Principle Component Analysis (PCA)

PCA was used to visualize the high-dimensional datasets. To eliminate null values, all counts in *HTSeq-count*-derived count table were increased by one. Transcript per Million (TPM) values were then calculated to correct for varying sequencing depth and gene length[21]. To reduce sensitivity to the uncertainty in TPM of low-expressing genes, PCA excluded genes for which the TPM averaged over all samples in the analysis was < 5. Finally, natural log of TPM values were used as input for PCA. For HUVEC dataset, PCA was first performed individually for each batch experiment. Based on the PCA results (**Figure 3.10-3.12**), several outliers were detected (based on separation from other biological replicates) and subsequently removed from further analysis: batch 1's control replicates 3, ALG-1001 replicates 5 and 6, batch 2's ALG-1001 replicates 6 and hypoxia replicates 6, and batch 3's ALG-1001+hypoxia replicates 6. PCA was performed again with all HUVEC samples excluding the outliers. For the PBMC dataset, PCA did not detect any obvious outliers (**Figure 3.33**).

EdgeR differential expression comparison

In the analysis of differential expression, it is beneficial to exclude genes that are not expressed (they reduce statistical power for identifying differential expression). Here, we retained expressed genes, defined as those with Count per Million (CPM) > 1 in at least four HUVEC samples or six PBMC samples (the smallest number of biological replicates from the same condition and batch experiment). CPM is calculated to normalize for sequencing depth [21]. Typically, around 13,000 genes remain after filtering. Next, GLM-based *edgeR* was used to perform the differential expression comparisons in the HUVEC study with batch variation removed from differential analysis[22]. Classic *edgeR* was used for the PBMC comparison. After statistical test, genes with False Discovery Rate (FDR) ≤ 0.05 were considered as differentially expressed (DE) and were further separated into up-regulated and down-regulated gene lists.

ERSSA

ERSSA was used to check whether the number of biological replicates used in the PBMC RNA-seq experiment is sufficient to identify the majority of DE genes[23]. Analysis was performed with absolute log2-fold-change cutoff at 0.5 with up to 50 subsamples generated at each replicate level. Since *ERSSA* currently does not support blocking, it was not used to test the HUVEC comparisons.

DE gene heatmap of log2-fold-change values

Heatmaps of TPM value were used to visualize the expression profile of DE genes. TPM values were first calculated from the unadjusted gene counts; then for each gene, TPM was

offset such that the minimum is zero and normalized so the maximum is one. DE genes were sorted row-wise from high to low *edgeR*-measured log2-fold-change values.

Goseq enrichment analysis of DE gene list

Functional analyses of the DE genes were performed with the *goseq* package, which mitigates gene length bias inherent to RNA-seq[24]. Enrichment was evaluated for both Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) biological pathways[25,26]. GO terms associated with genes were obtained from *biomart* while KEGG pathways were obtained from the KEGG REST server using the *keggrest* package[27,28]. Enrichment was performed separately for the up- and down-regulated gene lists. GO terms were considered enriched if FDR < 0.05. The GO enrichment results were visualized using *REVIGO,* a visualization tool that aggregates closely-related GO terms, with similarity level set to small and GO term size determined using the *UniPort Homo Sapiens* database[29]. For the particular cluster of GO terms that includes "angiogenesis" GO annotation, the cluster is represented by "angiogenesis" instead of the default selected GO annotation. Selected GO terms were labeled on the *REVIGO* plot based on our perception of their biological relevance. KEGG pathways were considered enriched if Benjamini-Hochberg adjusted p-value < 0.05 (p-value displayed in bar plots). In addition, "Human Diseases" KEGG pathways were removed to improve clarity.

Pathview

Pathview enables visualization of the fold changes across all expressed genes in any particular KEGG biological pathway[30]. *EdgeR* measured fold changes were used as input for *Pathview* visualization of selected pathways enriched in the PBMC analysis. For genes that share a common node on the KEGG pathway, the averaged fold change is displayed.

ISMARA

ISMARA was used to identify the transcription factors (TFs) that may be responsible for the observed profiles of gene expression across all comparisons[31]. Samples involved in specific comparisons were submitted together to *ISMARA* to obtain enrichment Z-scores. Analysis was focused on the top 50 most active motifs with Z-score > 1.0. Separately, all HUVEC samples were submitted together to obtain target gene activity profiles across all four conditions.

Data availability

Raw RNA-seq Fastq files will be available at Gene Expression Omnibus once the study has been published[32]. All enrichment tables will be made available at CaltechDATA[33].

Reverse transcription and quantitative polymerase chain reaction (qPCR)

Based on *edgeR* DE gene discovery, ALG-1001 regulated genes were selected for manual validation of expression among the HUVEC (six genes) and PBMC (three genes) samples. SuperScript IV Reverse Transcriptase (ThermoFisher) and oligo(dT)20 primer (ThermoFisher) were used to convert polyadenylated mRNA to cDNA. qPCR was performed using iTaq Universal SYBR Green Supermix (Biorad) on LightCycler 480

(Roche) with 45 amplification cycles. For HUVEC comparisons, six non-outlier replicates per condition from batch 1 and batch 2 (3 per batch experiment) were selected for qPCR. All PBMC samples were selected for qPCR. To enhance accuracy, four technical qPCR replicates of the same cDNA were generated and the median value selected for further analysis. Differences were calculated using the $\Delta\Delta C_t$ method versus *Actb*[34]. Primers used in this study were obtained from PrimerBank or designed using Primer-BLAST (**Table 2.1**)[35,36].

3.3 Results

WST-1 metabolic assay of HUVEC

ALG-1001 treatment in *in vivo* mouse models of retinal neovascularization generated strong anti-angiogenic effect (see **Chapter 1.2** for more detail, experiment performed by Campochiaro lab, Johns Hopkins University, unpublished). Here, we employed HUVECs to model human vascular endothelial cells and exposed them to an ALG-1001 concentration series to investigate the drug's anti-angiogenic property *in vitro*. First, cells seeded at low density were exposed to the drug for 2 days before measurement of cellular metabolic level as an indicator of cell density. During the incubation period, normal HUVEC growth media was supplied to promote active cell proliferation as during angiogenesis. Additionally, a separate set of samples were exposed to 24 hours of hypoxia to stimulate endothelial cells in a more disease-relevant environment (treatments described in **Figure 3.1**).

At end of the incubation period, metabolic assay shows HUVECs exhibited an ALG-1001 dose-dependent drop in metabolic activity (**Figure 3.8**). At room air oxygen tension (**Figure 3.8A**), a small reduction is already measurable at the lowest drug concentration tested (10μM) and continues to only around 30% of the control metabolic level at the highest drug concentration (5mM). The measurable decline is less apparent in the hypoxia treatment condition with only two of the highest concentrations tested (2.5 and 5.0mM) showing statistical significant reduction (**Figure 3.8B**). A higher biological variance was also observed with hypoxia co-treatment, which reduces our ability to accurately detect changes induced by ALG-1001.

Dead and live staining of HUVEC

To confirm the observation with metabolic assay is not due to drug-induced cell death, we performed dead and live staining of HUVECs at three ALG-1001 concentrations (500, 1000 and 2500μM). At both room air oxygen tension (**Figure 3.9A**) and 3% O_2 hypoxia (**Figure 3.9B**) conditions, we observed no obvious cell death or change in cell morphology in HUVECs after AGL1-1001 treatment. In addition, we found a noticeable drop in cell density at the highest drug concentration tested in both oxygen conditions (Control vs. 2500μM in **Figure 3.9A-B**). Combined with the metabolic assay findings, ALG-1001 appears to suppress vascular endothelial cell proliferation without causing cell death and morphology change.

PCA of HUVEC RNA-seq dataset

RNA-seq was used as an unbiased approach to explore the transcriptome regulation that may explain the metabolic assay and dead and live staining observations. For this study, four conditions were prepared by exposing HUVECs to vehicle or 500μM ALG-1001 in combination with room air oxygen tension or 24 hours of 3% O_2 hypoxia. To reduce batch effect and improve detection of the subtle effect of the drug on HUVEC transcriptome, the experiment was repeated three times to generate 18 biological replicates (6 replicates per study) for each condition.

Before performing differential expression analysis, PCA was first applied individually to each batch to visualize the RNA-seq samples in a low-dimensional space (**Figure 3.10-3.12**). In all three batch-specific PCAs, we see a clear separation between room air oxygen tension and hypoxia samples, suggesting the presence of a hypoxia-induced transcriptome shift. PCAs also show ALG-1001 treated samples appear to cluster close in space to control samples of the same oxygen level exposure, suggesting the drug effect is weaker than the effect of hypoxia. In terms of outlier detection, analysis of the top three principal components (PCs) revealed a total of six outliers (see **Figure 3.10-3.12** legend for specific samples) that are excluded from all subsequent analysis. Here, the outliers were identified base on strong separation from other replicates of the same condition, suggesting possible sample cross-contamination or other sample preparation issues.

After outlier removal, all non-outlier samples were visualized using PCA (**Figure 3.13**). In this analysis, the first three PCs captured a combined 48% of total variance. One feature that stood out is a clear separation on PC2 between batch 1 and batch 2-3 (**Figure 3.13B**).

Incidentally, the latter two batches were prepared around the same time while batch 1 was sequenced roughly half a year earlier. This observable batch effect suggests batch-specific transcriptome expression pattern exists and should be accounted for during differential expression analysis. Focusing on PC1, we also noticed separation between hypoxia and room air oxygen tension samples on a batch-specific basis, which serves as evidence of hypoxia-specific transcriptome regulation. On the other hand, separation between ALG-1001 treated samples and their respective control remains weak in all three PCs. In summary, PCA of HUVEC samples indicate: a) a strong batch effect exists in this dataset, b) there is an observable transcriptome shift by hypoxia treatment, and c) the expression regulation by ALG-1001 appears to be more subtle.

Hypoxia-associated transcriptome regulation in HUVEC

EdgeR was used to identify DE genes regulated by 24 hours of 3% O_2 hypoxia exposure in HUVECs (**Figure 3.14**). Between hypoxia and control conditions, we found 5376 genes differentially expressed, representing around 40% of the expressed transcriptome (**Figure 3.14A**). Of the 5376 DE genes, nearly all of them (5373/5376, 99%) are found to be moderately regulated with less than 1-fold change. This is in contrast to the large fold-changes detected in both the OIR mouse retina (**Chapter 2.3**) and ARPE-19 cell culture (**Chapter 4.3**) studies. Heatmap visualization of the DE gene expression indicates consistent perturbation across the experimental batches and fairly uniform expression among the biological replicates (**Figure 3.15**). Unlike our other studies, *ERSSA* was not used to confirm the appropriate sample size in this comparison since blocking (used in this study to remove

batch effect) is not currently supported by the software. Yet, the DE results suggest the number of replicates used is likely sufficient given the large number of small-fold-change DE genes discovered.

Next, we used *goseq* to identify the biological processes over-represented in the discovered DE genes. For this analysis, the DE genes were separated into two lists for investigation of up- and down-regulated biological processes (GO terms) and pathways (KEGG pathways). With the up-regulated genes, *goseq* found GO biological processes associated with known hypoxia-related mechanisms such as angiogenesis (*e.g.* "angiogenesis", "extracellular matrix organization"), inflammation (*e.g.* "immune system process", "cytokine production") and cell growth/dead (*e.g.* "cell proliferation", "cell death") (**Figure 3.16A**)[37]. In the enriched "angiogenesis" GO biological process, we found 125 associated genes, consisting of 39 positive regulator and 29 negative regulator of angiogenesis. VEGFA, which encodes for the potent angiogenic growth factor VEGF[38], was also found to be up-regulated. With the down-regulated genes, biological functions enriched include ones related to cell cycle (*e.g.* "cell cycle", "DNA replication") and metabolism (*e.g.* "heterocycle metabolism", "metabolism"), both of which are known to be regulated by hypoxia exposure[39,40].

KEGG pathway enrichment provides additional evidence and provide several new insights. With up-regulated genes, both angiogenesis- and inflammation-related pathways were enriched (**Figure 3.17A**). Out of the six pathways discovered, three ("ECM-receptor interaction", "Cell adhesion molecules", and "focal adhesion") are related to cell adhesion, which is actively involved in enabling vascular endothelial cell migration during

angiogenesis[41]. "Leukocyte transendothelial migration" is also enriched and is responsible for initiation of immune cell infiltration during inflammation[37]. Other enrichment results suggest regulation of energy-intensive translational processes in response to hypoxia: "ribosome" genes were found to be up-regulated while three of the four pathways enriched with down-regulated genes ("ribosome biogenesis in eukaryotes", "Aminoacyl-tRNA biosynthesis" and "RNA transport") are involved in various aspects of translation (**Figure 3.17A-B**). Lastly, "steroid biosynthesis" and "Fanconi anemia pathway" were also enriched, but their adjusted p-values are very weak and their relationship with hypoxia is unclear.

Transcriptome regulation by ALG-1001 exposure in HUVEC

In contrast to the transcriptome regulation by hypoxia, the effect that ALG-1001 has on HUVEC transcriptome is relatively small. Here, we performed two comparisons to identify the genes regulated by the drug at both room air oxygen tension and with hypoxia co-treatment conditions. In the two comparisons, we found 1470 (room air oxygen tension) and 922 (hypoxia) DE genes, respectively (**Figure 3.18**). Heatmap visualization of the DE gene expression shows consistent expression pattern across the three batches with some heterogeneity among the biological replicates (**Figure 3.19-20**). *ERSSA* was not applied to these comparisons due to lack of blocking capability in existing software.

Functional enrichment of the regulated genes generated both expected as well as surprising results. With the DE genes activated by ALG-1001 at room air oxygen tension, we observed enrichment of GO biological processes related to angiogenesis ("angiogenesis", "circulatory

system development", "extracellular matrix organization") and inflammation ("inflammatory response") (**Figure 3.21A**). Among the up-regulated genes, 48 are found to be involved with "angiogenesis" with 13 positive and 12 negative regulators. Interestingly, we also found 32 "angiogenesis" genes down-regulated including the positive regulator – VEGFA. However, GO enrichment with the down-regulated DE genes mainly identified several GO biological processes related to metabolism and translation (**Figure 3.21B**). Together, we found evidence of a selective up-regulation of angiogenesis and inflammation genes after ALG-1001 treatment at room air oxygen tension.

In comparison, the regulation by ALG-1001 with hypoxia co-treatment shows down-regulation of many processes activated by hypoxia (**Figure 3.22B**). Several noteworthy GO biological processes up-regulated by hypoxia and down-regulated by ALG-1001 co-treatment include "angiogenesis", "cell proliferation", "cell death", "extracellular matrix organization", and "actin cytoskeleton organization". In fact, more than 62% (135/218) of these enriched biological processes were previously enriched with hypoxia up-regulated DE genes. Complementary analysis of cellular component and molecular function GO annotations found strong evidence that the down-regulated genes function in the extracellular matrix and are involved in cell-adhesion (**Figure 3.44**). In addition, around 33% (303/922) of the ALG-1001-regulated genes in this comparison show opposite regulatory pattern compare to hypoxia treatment, supporting the observation that ALG-1001 moderated the effect of the low oxygen tension (**Table 3.2**, versus 21%, 196/922 that show same regulatory pattern). Taking one step further, we also found a weak but inverse relationship (r=-0.160)

between the fold change of the hypoxia-regulated genes and their fold change with ALG-1001 co-treatment (**Figure 3.23**). It is worth mentioning that among the ALG-1001 down-regulated genes, 38 are "angiogenesis" genes with 13 positive and 6 negative regulators. VEGFA, previously found to be up-regulated by hypoxia, was also suppressed by ALG-1001 during hypoxia co-treatment. In summary, with the cells stimulated in a disease-relevant hypoxia state, ALG-1001 appears to moderate some of the hypoxia-activated transcriptome regulation.

Since "angiogenesis" GO biological process was enriched in all three comparisons, we set out to investigate the associated DE genes in more detail (**Figure 3.25**). Across the three comparisons, we found hypoxia treatment disproportionally increased "angiogenesis" gene expression (125/190 or 66% of regulated "angiogenesis" genes are elevated), while ALG-1001 without hypoxia appeared to be more moderate (48/80 or 60% of regulated "angiogenesis" genes are elevated) and ALG-1001 under hypoxia condition with the reverse trend (26/64 or 41% of regulated "angiogenesis" genes are elevated). In terms of the fold change, the regulatory pattern of ALG-1001-modulated "angiogenesis" genes is quite consistent regardless of the hypoxia co-treatment (**Figure 3.25B-C**). Additionally, among the 354 genes regulated by ALG-1001 in both comparisons, 100% of them have the same regulatory direction (**Table 3.3**). On the other hand, expression pattern of hypoxia-regulated "angiogenesis" genes are more heterogeneous among the three comparisons, with the highest consistency still between the two ALG-1001 comparisons (**Figure 3.25A**). These

observations show in contrary to the *goseq* results, specific fold change of ALG-1001-regulated "angiogenesis" genes are consistent across both oxygen conditions.

Limited KEGG pathways were enriched with ALG-1001 regulated genes (**Figure 3.24**). In both comparisons, we found the "steroid biosynthesis" pathway enriched with ALG-1001 up-regulated genes and "aminoacyl-tRNA biosynthesis" pathway enriched with ALG-1001 down-regulated genes. Out of the two, we found proteins in the "aminoacyl-tRNA biosynthesis" pathway have alternative functions in regulating angiogenesis. Specifically, the tRNA synthases encoded by YARS, WARS, TARS (down-regulated by ALG-1001 in both comparisons) have alterative functions in activating angiogenesis, immune cell migration, endothelial cell detachment and migration[42].

Transcription factors enriched from HUVEC RNA-seq dataset

ISMARA was used to identify the TFs that are likely responsible for the observed gene expression regulation in HUVECs. For the TFs predicted to be involved in regulating hypoxia response, many of the top 50 enriched are relevant to the biology of hypoxia (**Figure 3.26, Table 3.4**). The associated Z-scores of these TFs are relatively high, which suggests an abundance of evidence for their involvement in hypoxia transcriptome regulation. Notable TFs with elevated target expression include those involved in regulating hypoxia-related processes such as cell growth/death (SMAD4, SMAD1, WRNIP1, KLF4, PLAGL1, BCL3, and ERG), angiogenesis (GATA3), metabolism (ESRRA), and hypoxia-response (EPAS1) (**Figure 3.27**). Similarly, many of the TFs with down-regulated target genes are found to be

involved in regulating cell growth/death (HCFC1, TAF1, SIN3A, MYB, ATF2, and SP1), angiogenesis (E2F7, E2F1, E2F3, TFDP1, and ZNF384), metabolism (NRF1, AHR), inflammation (STAT5A), and hypoxia-response (ARNT2, CREB1) (**Figure 3.28**).

Consistent with the OIR mouse dataset (**Chapter 2.3**), the enrichment of TFs regulated by ALG-1001 is less statistically significant compared to hypoxia. For both oxygen tension conditions, we found less than 20 TF motifs with Z-score>1.0, which is typically used as the statistical cutoff for *ISMARA* (**Figure 3.29, 3.31, Table 3.5-3.6**). When analysis was narrowed to those TFs that passed the statistical cutoff, we found many of them are in fact related to processes regulated by hypoxia. At room air oxygen tension, we observed enrichment of TFs regulating angiogenesis (GATA3, ID4), cell growth/death (SMAD4, WRNIP1, ERG, FOS, TAF1, and KLF8) and inflammation (CEBPB) (**Figure 3.30**). At hypoxia condition, only 12 TFs are enriched with 6 involved in angiogenesis (ZNF384), cell growth/death (SP1, MNT, MYC, and KLF4) and inflammation (CEBPB) (**Figure 3.32**). Intrigued by functional analysis that showed ALG-1001 is able to moderate hypoxia transcriptome regulation, we looked for the same pattern in the *ISMARA* results. Interestingly, we observed that out of the 18 TF motifs that are found on both the list of top 50 enriched TFs (**Table 3.6**, labeled in bold), 15 of them showed a reverse target gene activity. In addition, many of these 15 regulators are related to hypoxia biology, including SMAD1, SMAD4, KLF4, PLAGL1, ZNF384, ESRRA, E2F3, and SP1.

<u>Transcriptome regulation by ALG-1001 exposure in PBMC</u>

Immune cells' involvement in the pathophysiology of retinal diseases has been extensively reported in prior literature[43,44]. In OIR dataset (**Chapter 2.3**), we found disease-related hypoxia stimulation activated both angiogenesis and immune gene expression, suggesting modulation of immune system together with neovascularization. After ALG-1001 treatment in the neovascularizing retina, we also discovered evidence that hypoxia-activated immune genes are suppressed by the drug treatment. To further investigate ALG-1001's involvement in regulating immune cells, we treated human donor PBMCs with ALG-1001 and probed the transcriptome regulation after drug treatment. For this study, six replicates of PBMC treated with vehicle or ALG-1001 were sequenced with RNA-seq. PCA of the expression profile indicates weak separation on PC1 (accounting for 31% of total variance) between the treatment conditions (**Figure 3.33**). *EdgeR* comparison identified 1204 DE genes regulated by ALG-1001 with few by more than 1-fold (**Figure 3.34A**). Heatmap of the DE genes shows moderate amount of expression variation between biological replicates, suggesting some amount of inconsistency between the replicates (**Figure 3.35**). We also employed *ERSSA* to test the detection power of the current sample size and the results suggest not all DE genes have been discovered (double digit percentage increase in DE gene discovery up to the full sample size, **Figure 3.34C-D**). As a result, the following functional analysis should be viewed as preliminary as more genes are likely regulated by ALG-1001.

Goseq was used to identify GO biological processes associated with up-regulated and down-regulated genes. With up-regulated genes, we found enrichment of "lymphocyte activation", "regulation of immune system process", and "activation of immune response" processes.

"rRNA processing" was also strongly enriched, suggesting regulation of translational processes (**Figure 3.36A**). With down-regulated genes, we mainly found enrichment of immune-related processes such as "immune response", "leukocyte activation involved in immune response", "inflammatory response", and "leukocyte migration" (**Figure 3.36B**).

KEGG pathway enrichment of the DE genes are consistent with the GO biological process analysis. With up-regulated genes, we found enrichment of "ribosome" and "T cell receptor signaling" pathways (**Figure 3.37A**). Consistent with GO analysis, enrichment of "ribosome" pathway suggest modulation of translational processes by ALG-1001. On the other hand, enrichment of "T cell signaling pathway" is intriguing, as stimulation of the pathway activates T cells, leading to their proliferation, differentiation, and immune modulation. Visualization of the "T cell signaling pathway" gene expression shows the expression profile is mixed, with predominant up-regulation of genes that encode for proteins that initiate the signaling process and down-regulation of genes down-stream in the pathway (**Figure 3.38**). With down-regulated genes, we see enrichment of many pro-inflammatory pathways, including "TNF", "C-type lectin receptor", "IL-17", and "Toll-like receptor" signaling pathways (**Figure 3.37B**). In particular, TNF signaling pathway is a potent pro-inflammatory process that has reported to be activated in retinal diseases[45–47]. Visualization of the pathway gene expression shows down-regulation of many pathway components, including genes that are transcriptionally regulated by the pathway (**Figure 3.39**).

Using *ISMARA*, We found selected TFs that are likely involved in ALG-1001 treatment response and are biologically relevant to retinal disease pathophysiology (**Figure 3.40-3.41,**

Table 3.7). Specifically, some of the enriched transcriptional regulators are involved in processes such as angiogenesis (E2F7, E2F1, E2F3, and ZNF384), cell growth/death (FOXO3, FOSL2, PLAGL1, SMAD1, and SMAD4) and inflammation (NFKB2, IKZF1, IRF4, and IRF5). Among the TFs enriched, NFKB2 is particularly interesting as it is a member of the NF-κB family of transcription factors that, once activated, lead to expression of pro-inflammatory cytokines, chemokines and adhesion molecules[48]. NF-κB has also previously been found to be activated in mouse models of retinal diseases and in retinal cells of diabetic patients[49–51]. Here, we found that expression of NFKB2 target genes are down-regulated by ALG-1001, indicating possible suppression of pro-inflammatory gene expression by the drug through NF-κB (**Figure 3.41B**).

qPCR (HUVEC and PBMC)

qPCR was used to validate six selected genes found to be regulated by ALG-1001 in HUVEC (**Figure 3.42**). Compared to the RNA-seq measurements, some of the qPCR gene expression quantification's standard deviation are noticeably higher. In addition, a few genes found by RNA-seq to be up-regulated in hypoxia condition appeared to be unchanged with qPCR measurement. The discrepancy is likely due to the fairly small expression change observed, which may be within the range of qPCR technical error. Overall, a large fraction of regulatory patterns are consistent between the two results despite technical limitation.

For the PBMC dataset, we aimed to validate the RNA-seq result with three immune-relevant genes (CCL2, CSF1, and RELB) using qPCR. For each gene, RNA-seq showed ALG-1001

weakly down-regulated their gene expression (**Figure 3.43A**), while the corresponding qPCR measurements largely confirm these results except for the expression pattern of RELB (**Figure 3.43B**). For this gene, the expression change is likely too small to be accurately detected by qPCR, as evident by the relatively large standard deviation. In short, we were able to observe the same expression pattern for two of the three genes tested while the expression change associated with the RELB appears to be too small to be accurately qualified by qPCR.

3.4 Discussion

Endothelial transcriptome modulation by hypoxia and ALG-1001

Diseases of the retina affect millions of patients worldwide with few effective treatment options available[12,13,52]. With the current antibody-based standard of care ineffective in double digit percentage of patients, new solutions are sorely needed[6,7]. In recent years, the development of ALG-1001 as a potential therapy has produced promising clinical trial results as well as positive data in mouse (see **Chapter 1.2 and 2.1** for more detail)[10,53]. Here, we applied RNA-seq to study the therapeutic effect of the drug in two human primary cell culture systems that represent cells relevant to disease pathophysiology. In particular, low passage HUVECs were used to study the drug's effect on proliferating endothelial cells that play an essential role in disease-associated retinal neovascularization[54]. Freshly collected PBMCs were also tested to identify ALG-1001's effect on immune cells that generally promote a cytotoxic inflammatory environment in diseased retinal tissue[14–16]. Through these

approaches, we found weak, but specific regulation by the drug on angiogenesis and inflammation.

Using proliferating HUVECs, we set out to validate the anti-angiogenic effect observed in the OIR mouse model. Employing metabolic activity measurement as a proxy for cell density, we observed a dose-dependent suppression of endothelial cell proliferation by ALG-1001 (**Figure 3.8**). In addition, fluorescent staining confirmed this effect is through inhibition of cell proliferation, instead of by inducing cell death as previously observed with structurally similar RGD peptide (**Figure 3.9**)[55]. Collectively, metabolic and imaging assays suggest ALG-1001 can suppress endothelial proliferation, even in a highly angiogenic cell culture environment where cells are induced to undergo rapid cell division.

Next, RNA-seq was used as an unbiased approach to study HUVEC's transcriptome response to disease-relevant hypoxia exposure and ALG-1001 treatment. After 24 hours of 3% O_2 low oxygen exposure, differential expression analysis found more than 5000 genes regulated by the treatment (**Figure 3.14**). However, compared to the magnitude of the hypoxia-induced expression change observed in OIR mouse retina (**Chapter 2.3**) and ARPE-19 cells (**Chapter 4.3**), the fold changes are noticeably smaller in HUVECs. We suspect this is likely due to different cell microenvironments among the three models. In both OIR mouse model and ARPE-19 cell culture experiments, the retinal cells reside in a low proliferating state before hypoxia exposure. On the other hand, the HUVEC experiments were conducted in nutrient-rich growth media that promotes fast proliferation of the vascular endothelial cells. Thus, it is likely that HUVEC transcriptome is already in a state that sustains rapid cell

proliferation, leaving little room for hypoxia to induce further transcriptome change in the direction of promoting cell division.

While the magnitude of expression changes is relatively small, functional analysis of the roughly five thousand DE genes enriched many biological processes relevant to hypoxia biology. In particular, GO enrichment with hypoxia up-regulated genes revealed biological processes related to angiogenesis, cell growth/death, inflammation and hypoxia response (**Figure 3.16A**)[56,57]. Similarly, down-regulated genes are over-represented with those related to hypoxia response such as cell cycle and metabolism (**Figure 3.16B**)[39,40]. In agreement, KEGG pathway enrichment found up-regulation of genes involved in cell adhesion (an important component of angiogenesis) and immune cell transendothelial migration (**Figure 3.17A**)[37,41]. However, unlike biological process analysis, KEGG pathway enrichment of hypoxia down-regulated genes identified three pathways related to translation, but not those related to cell cycle and metabolism (**Figure 3.17B**). In relation to translation, hypoxia is known to suppress translation due to a state of cellular energy shortage[58].

Compared to the hypoxia-induced transcriptome change, ALG-1001's effect on gene expression is relatively small. After 48 hours of ALG-1001 treatment, we found 1470 and 922 DE genes at room air oxygen tension and hypoxia conditions, respectively (**Figure 3.18**). Of the genes regulated by ALG-1001 at room air oxygen tension, we are surprised to find over-representation of several GO biological processes that were previously found to be associated with hypoxia response (**Figure 3.21**). In particular, there are several that suggest selective up-regulation of angiogenesis genes by ALG-1001. In terms of fold change, we

found a positive correlation (r=0.407) among hypoxia-regulated "angiogenesis" genes between hypoxia-response and ALG-1001 response at room air oxygen tension (**Figure 3.25A**). Taken together, these results suggest, on face value, that ALG-1001 may be activating angiogenesis instead of suppressing it.

It is worth noting that there are also "angiogenesis" genes down-regulated by ALG-1001 that are up-regulated by hypoxia. In fact, we noticed several important angiogenesis genes that are up-regulated by hypoxia and suppressed by ALG-1001 at both oxygen treatment conditions. Four notable genes that exhibit this pattern include VEGFA and ANXA3, both of which are potent pro-angiogenesis proteins[38,59]; TGFB2, which was reported to suppress angiogenesis[60]; and TGFBI, which plays a role in cell adheison[61]. Additionally, both VEGFA and ANXA3 were also observed to be down-regulated by ALG-1001 in the OIR mouse retina (**Chapter 2**), indicating possible regulation of angiogenesis through modulation of these key players.

Interestingly, the transcriptome regulation by ALG-1001 with hypoxia co-treatment tells a different story. Under hypoxic condition, we detected an inverse relationship between hypoxia-induced gene expression and after co-treatment with ALG-1001 (Pearson's r=-0.160, **Figure 3.23**), which suggest ALG-1001 moderated some of transcriptome response to hypoxia stimulation. Additionally, more than half of the GO biological processes enriched with ALG-1001-suppressed genes (*e.g.* "angiogenesis", "cell death", "cell proliferation", and "extracellular matrix organization") are enriched with hypoxia up-regulated genes (**Figure 3.22**). KEGG pathway enrichment supports these findings with enrichment of "aminoacyl-

tRNA biosynthesis" pathway among ALG-1001-suppressed genes (**Figure 3.24**). Specifically, we found three tRNA synthases (YARS, WARS, and TARS) down-regulated by ALG-1001 also have alternative function in modulating angiogenesis[42]. For instance, tyrosyl-tRNA synthase (YARS) can be secreted and cleaved into cytokine fragments with pro-angiogenic effect, while threonyl-tRNA synthase (TARS) is secreted to stimulate vessel migration, patterning and maturation. On the other hand, Tryptophanyl-tRNA synthase (WARS) appears to inhibit angiogenesis, where its cleaved form can be secreted to disrupt endothelial cell attachment and angiogenic signaling molecule. Since tRNA synthases function in both translation and angiogenesis regulation, it is unclear from gene expression alone whether they are actually involved in ALG-1001's anti-angiogenic effect. Additional biochemistry and genetics experiments could help clarify their involvement.

A key component of drug mechanism of action study is identification of both the drug target and the modulated down-stream effector proteins. With RNA-seq data, we are unable to directly identify either of these components, but we can use the gene expression profile to infer TFs that are likely involved in driving the gene regulation. For this analysis, we used *ISMARA* to identify the factors that are predicted to participate in hypoxia response and after ALG-1001 treatment. When applied to study hypoxia, the algorithm enriched many TFs that are known regulator of hypoxia-regulated processes such as cell growth/death, angiogenesis, metabolism, inflammation, and hypoxia-response (**Figure 3.26**). In terms of ALG-1001 transcriptome modulation, limited number of TFs were enriched with statistical significance (**Figure 3.29, 3.31**). Interestingly, enriched factors include those involved in angiogenesis,

cell growth/death, and inflammation processes. Additionally, ALG-1001 treatment reversed target gene activity for several TFs that are regulated by hypoxia (*e.g.* SMAD family proteins, SP1 and KLF4), suggesting these TFs may play a role as down-stream effector of ALG-1001. Intriguingly, these factors are also involved in regulating a wide range of biological activities, including cell growth, adhesion, apoptosis, development, metabolism, immune system and morphogenesis[62–66]. Although RNA-seq data does not provide concrete evidence that these TFs are involved in ALG-1001's therapeutic effect, our findings warrant further investigation to confirm their roles in the drug's mechanism of action.

Immune cell gene expression modulated by ALG-1001

In a second study, fresh human donor PBMCs were treated with low dosage of ALG-1001 for 8 hours and then probed with RNA-seq to identify the genes regulated. Comparable to the HUVEC results, we found around 1200 DE genes, many with small fold changes (**Figure 3.34**). Interestingly, functional analysis of the enriched genes suggest regulation of multiple immune processes by ALG-1001 (**Figure 3.36, 3.37**). Using KEGG pathway enrichment, we found over-representation of "TNF signaling pathway" genes in the down-regulated gene list (**Figure 3.37B**). Visualization of the changes in the pathway shows that majority of the pathway is weakly down-regulated by ALG-1001 (**Figure 3.39**). Intriguingly, we also enriched "T cell receptor signaling pathway" with up-regulated genes. T cell receptor signaling is activated after successful antigen binding and its activation suggest possible T cell receptor recognition of ALG-1001, leading to initiation of T cell activity (**Figure 3.37A, 3.38**). In terms of *ISMARA* TF enrichment, we found: a) enrichment of NFKB2, a member

of the pro-inflammatory NF-κB protein family and b) identification of SMAD4 and SNF384, which were also enriched in the HUVEC experiment. In particular, ALG-1001 suppressed these two factor's target gene activity in both models, indicating a common predicted regulatory effect.

While these results are promising, the PBMC study was repeated only once with PBMCs collected from a single donor. As a result, these findings should be considered as preliminary as additional samples are needed for more concrete conclusions (suggested by *ERSSA*, **Figure 3.34C-D**). Additionally, PBMCs are composed of multiple immune cell types and bulk RNA-seq cannot easily discern the effect of ALG-1001 on individual cell types. More studies should be carried out with purified samples of specific cell types, with monocytes as a priority due to their known involvement in retinal diseases[67].

Limitation of *in vitro* cell culture

In general, results from *in vitro* cell culture should be further validated *in vivo*. In cell culture, the cells are removed from their *in vivo* tissue microenvironment, which can alter their response to drug treatment[68]. Indeed, as in most cell culture studies, the culture environment employed in our studies are artificial: a) high level of oxygen tension (typical room air oxygen percentage of around 21% vs. around 5% in normal tissue[69]), b) nutrient-rich growth media (induce rapid cell proliferation with growth factor levels much higher than those found in tissue), c) cells grown in monolayer without interaction with other cell types. Additionally, this study have the added challenge of trying to mimic the diseased microenvironment in cell

culture, which is very challenging. Overall, our results from cell culture studies should be viewed in context of these limitations, such that any observations *in vitro* need to be validated with the appropriate animal model.

3.5 Conclusions

Retinal diseases are currently untreatable for millions of patients worldwide. ALG-1001 is an oligopeptide drug that is currently under clinical investigation for treatment of these blinding eye diseases. In this study, we used human primary cell culture models of vascular endothelial and immune cells to study the drug's therapeutic effect, through analysis of the treatment-associated transcriptome regulation. In endothelial cells, we found ALG-1001 modulated expression of angiogenesis genes activated by disease-relevant hypoxia exposure. In immune cells, we found ALG-1001 treatment down-regulated pro-inflammatory TNF signaling pathway, potentially through modulation of NF-κB. Altogether, our results provided evidence of cell type-specific regulation by ALG-1001, particularly in suppressing disease-worsening angiogenesis and inflammation. It is also important to view these results in context of the limitations of cell culture models, which always require further validation with animal studies.

3.6 Figures and Tables

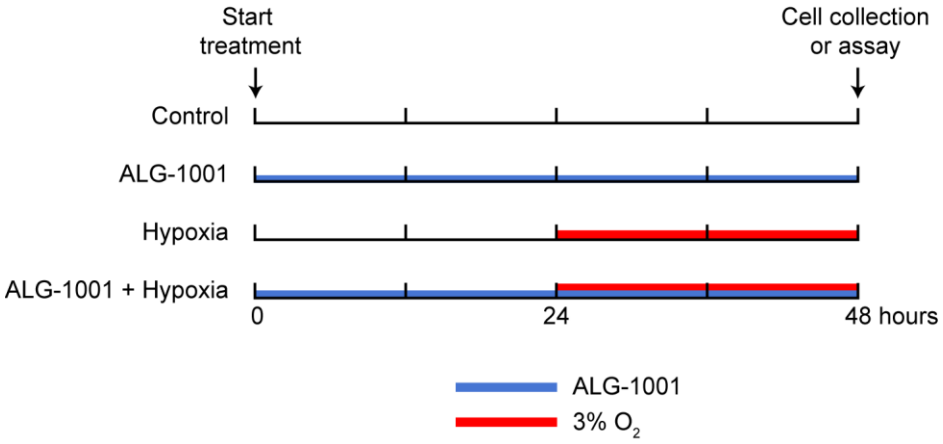

Figure 3.1 HUVEC treatment design
HUVECs were exposed to vehicle (empty) or ALG-1001 (blue bar) at experiment-specific concentration for 48 hours and with or without 3% O_2 hypoxia (red bar) for 24 hours before assay. This treatment scheme is applied to prepare WST-1 assay, Dead/live staining and RNA-seq samples.

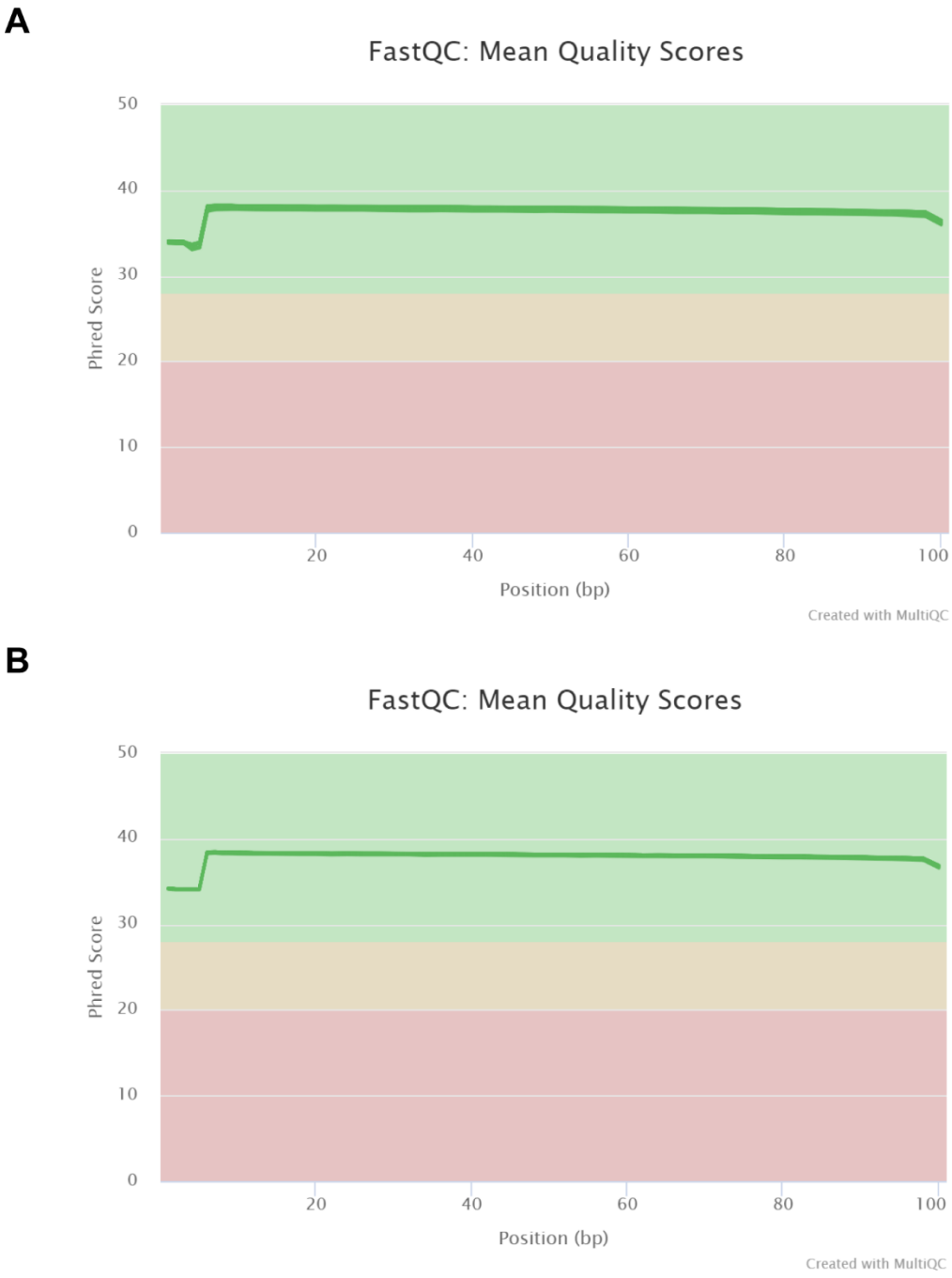

Figure 3.2 *FastQC* mean sequencing quality score – part 1
MultiQC plots the average sequencing quality Phred score at each base in the reads generated for each sample. Each line detonates a sample, with majority of them overlapping throughout the entire 100 base pair region. Phred score plot indicates good sequence quality across all samples. **A.** HUVEC study, batch 1 samples **B.** HUVEC study, batch 2 samples.

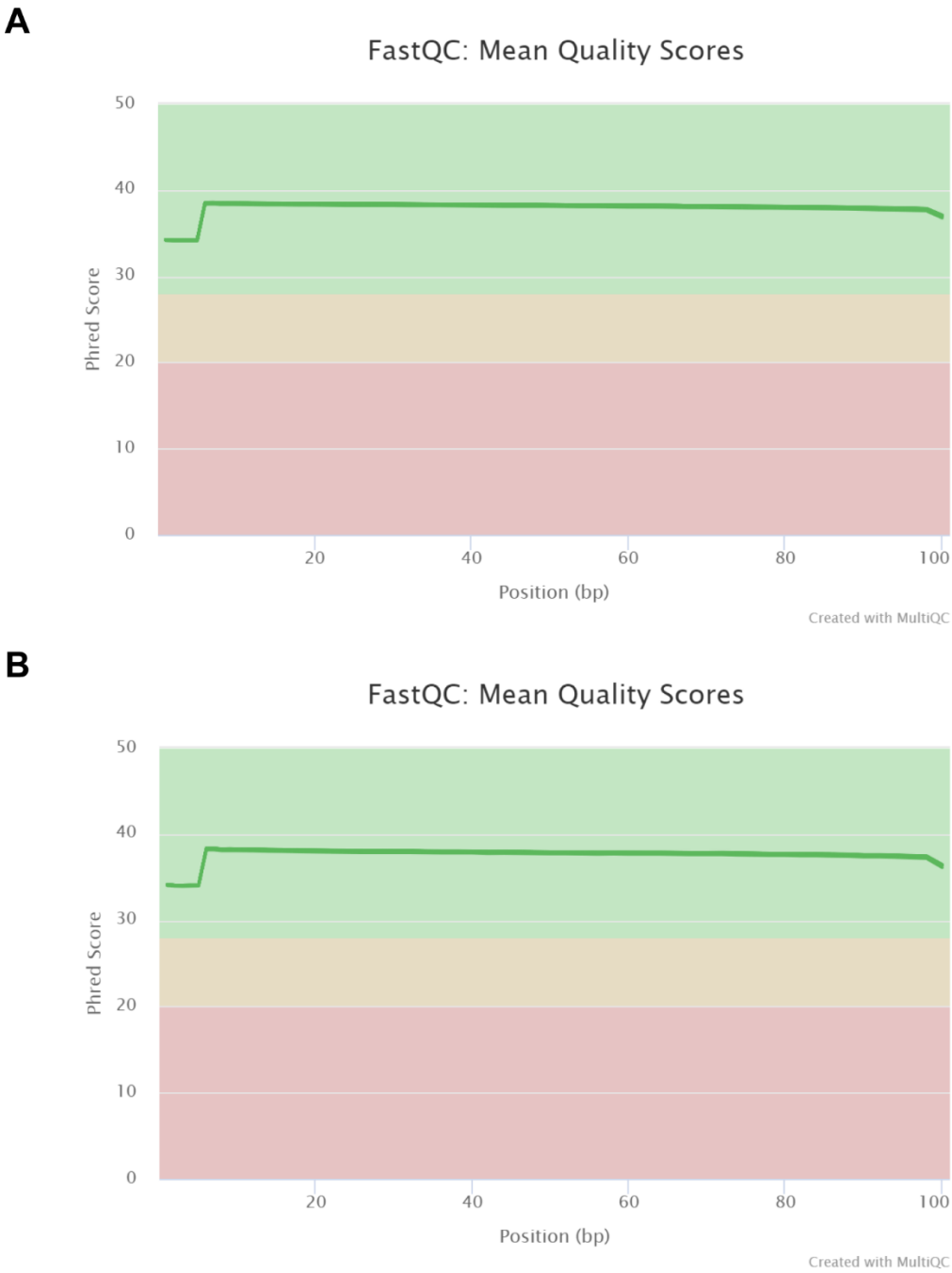

Figure 3.3 *FastQC* mean sequencing quality score – part 2
MultiQC plots the average sequencing quality Phred score at each base in the reads generated for each sample. Each line detonates a sample, with majority of them overlapping throughout the entire 100 base pair region. Phred score plot indicates good sequence quality across all samples. **A.** HUVEC study, batch 3 samples **B.** PBMC study samples.

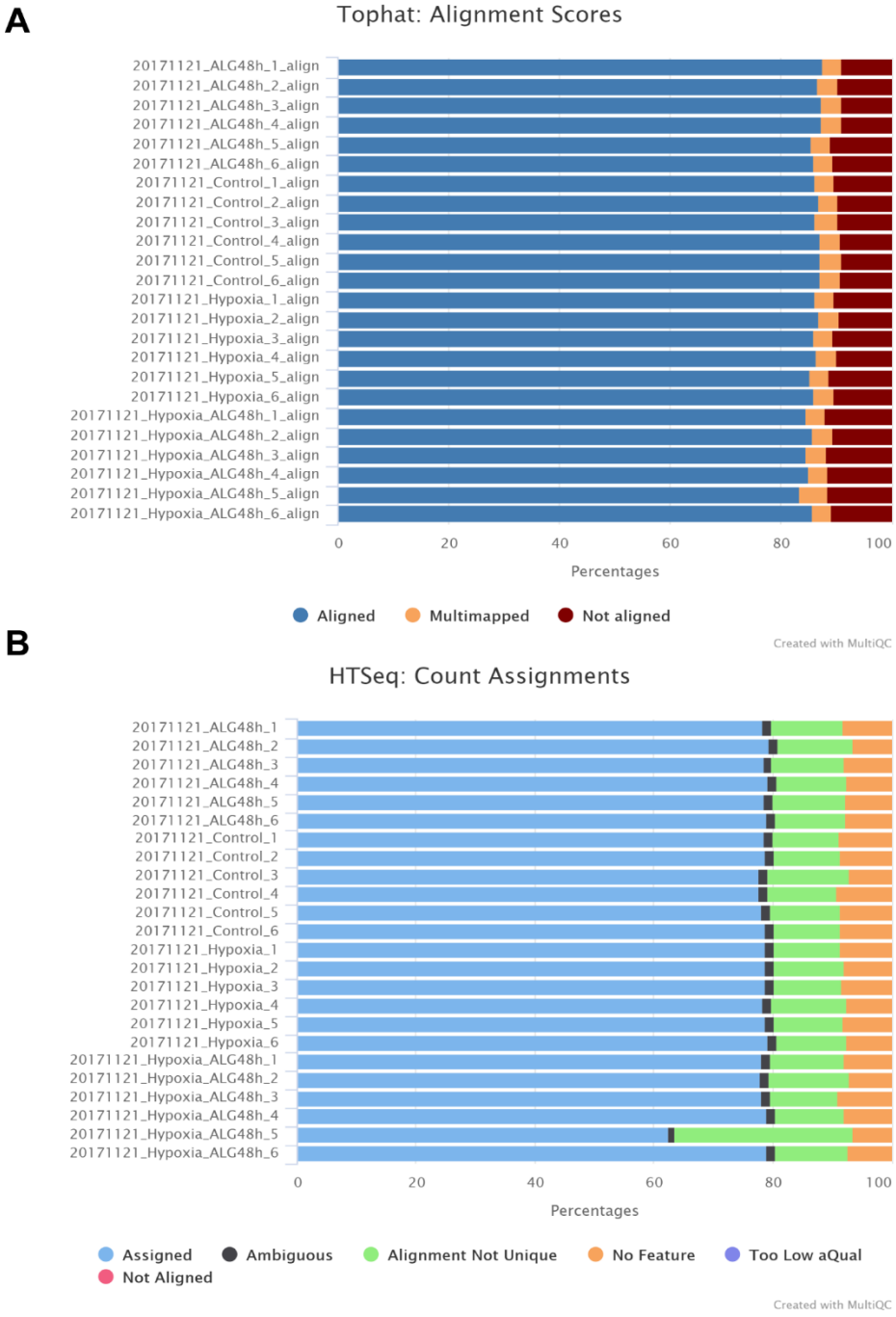

Figure 3.4 *Tophat2* and *HTSeq-count* statistics – HUVEC, batch 1 samples
A. At least 88% of reads were mapped in all samples by *Tophat2*. **B.** *HTSeq-count* uniquely assigned (blue bar) between 77-79% of a sample's aligned reads to a gene except replicate 5 of ALG-1001+hypoxia condition (62.5%).

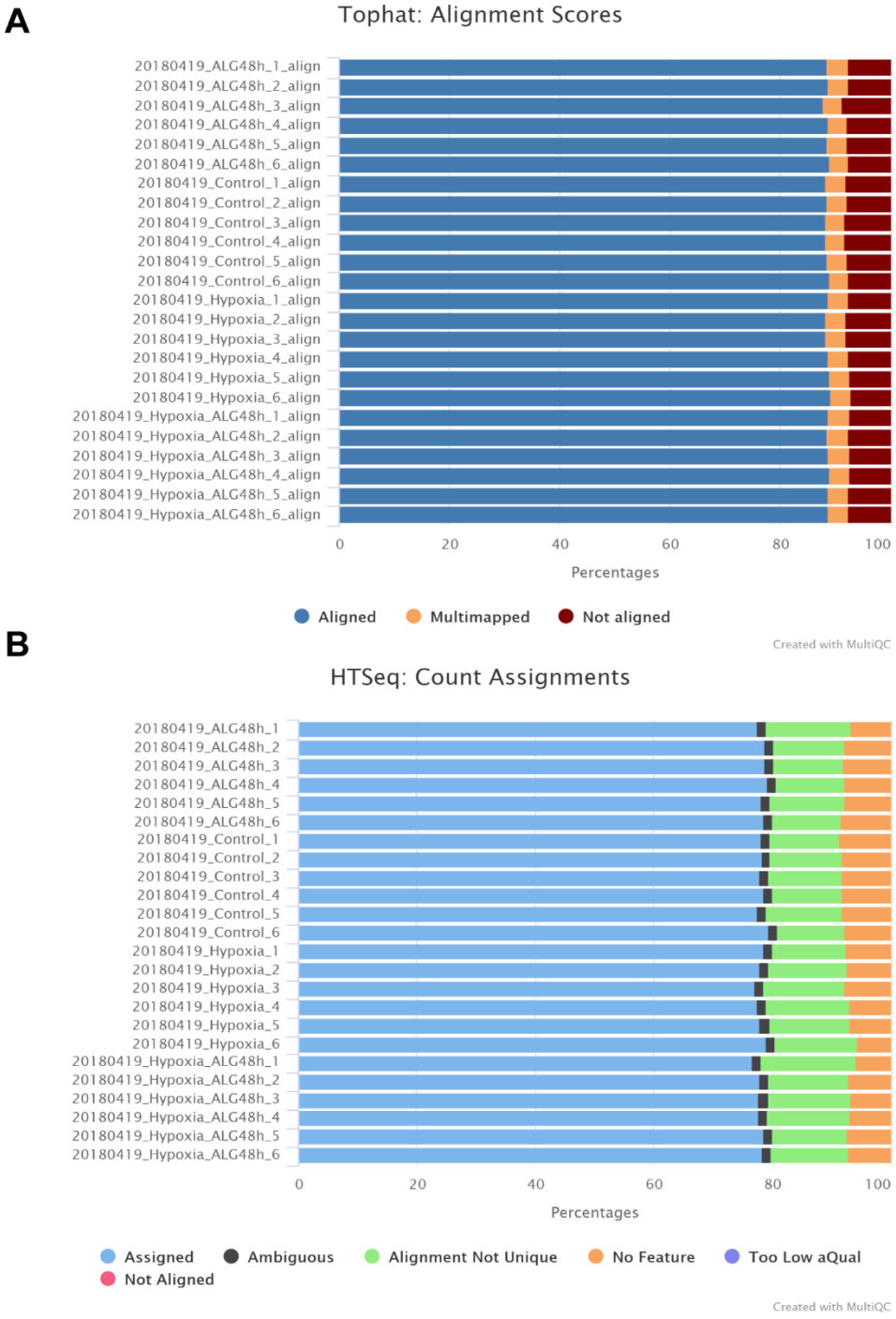

Figure 3.5 *Tophat2* and *HTSeq-count* statistics – HUVEC, batch 2 samples
A. At least 91% of reads were mapped in all samples by *Tophat2*. **B.** *HTSeq-count* uniquely assigned (blue bar) between 76-79% of a sample's aligned reads to a gene.

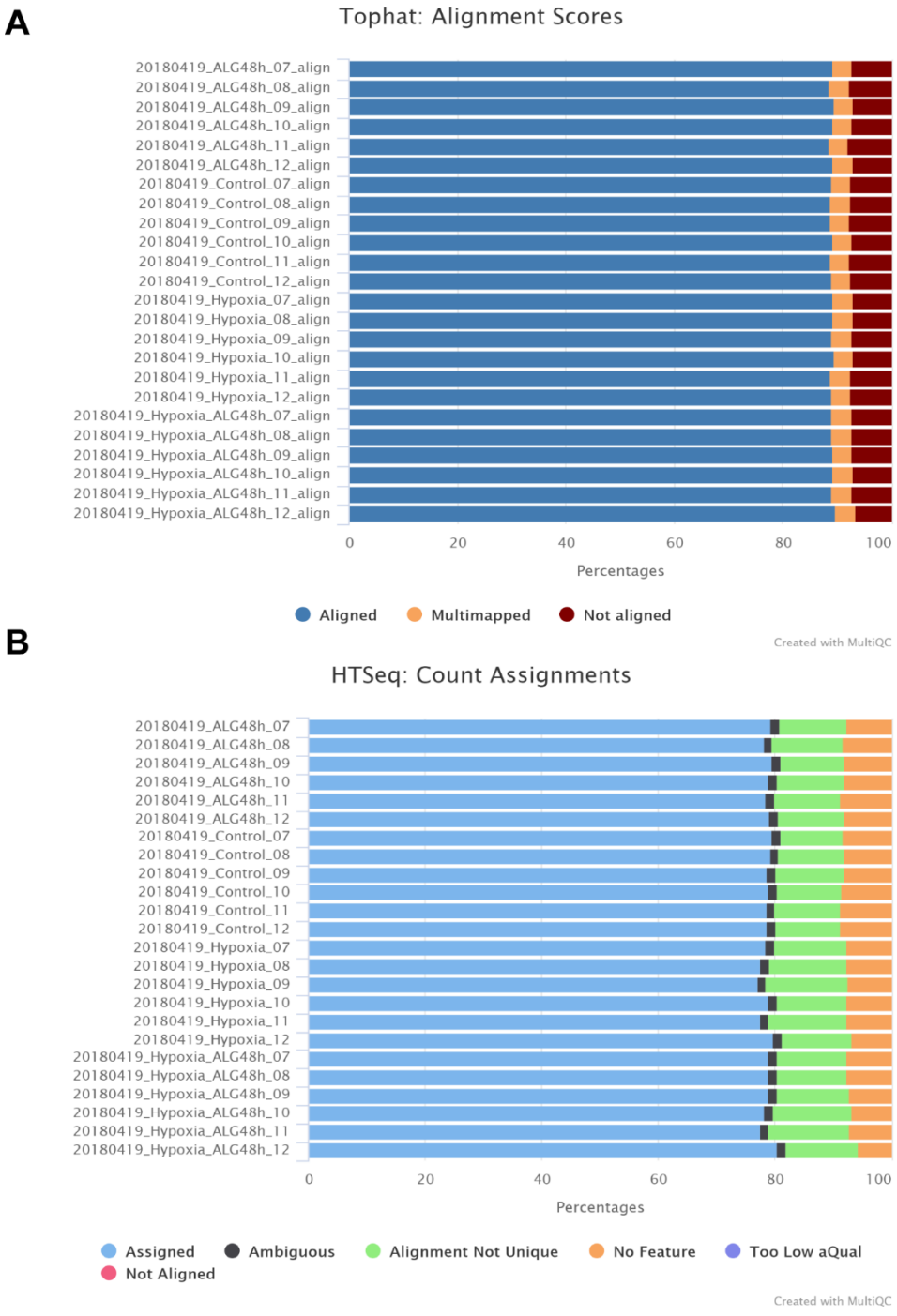

Figure 3.6 *Tophat2* and *HTSeq-count* statistics – HUVEC, batch 3 samples
A. At least 92% of reads were mapped in all samples by *Tophat2*. **B.** *HTSeq-count* uniquely assigned (blue bar) between 77-80% of a sample's aligned reads to a gene.

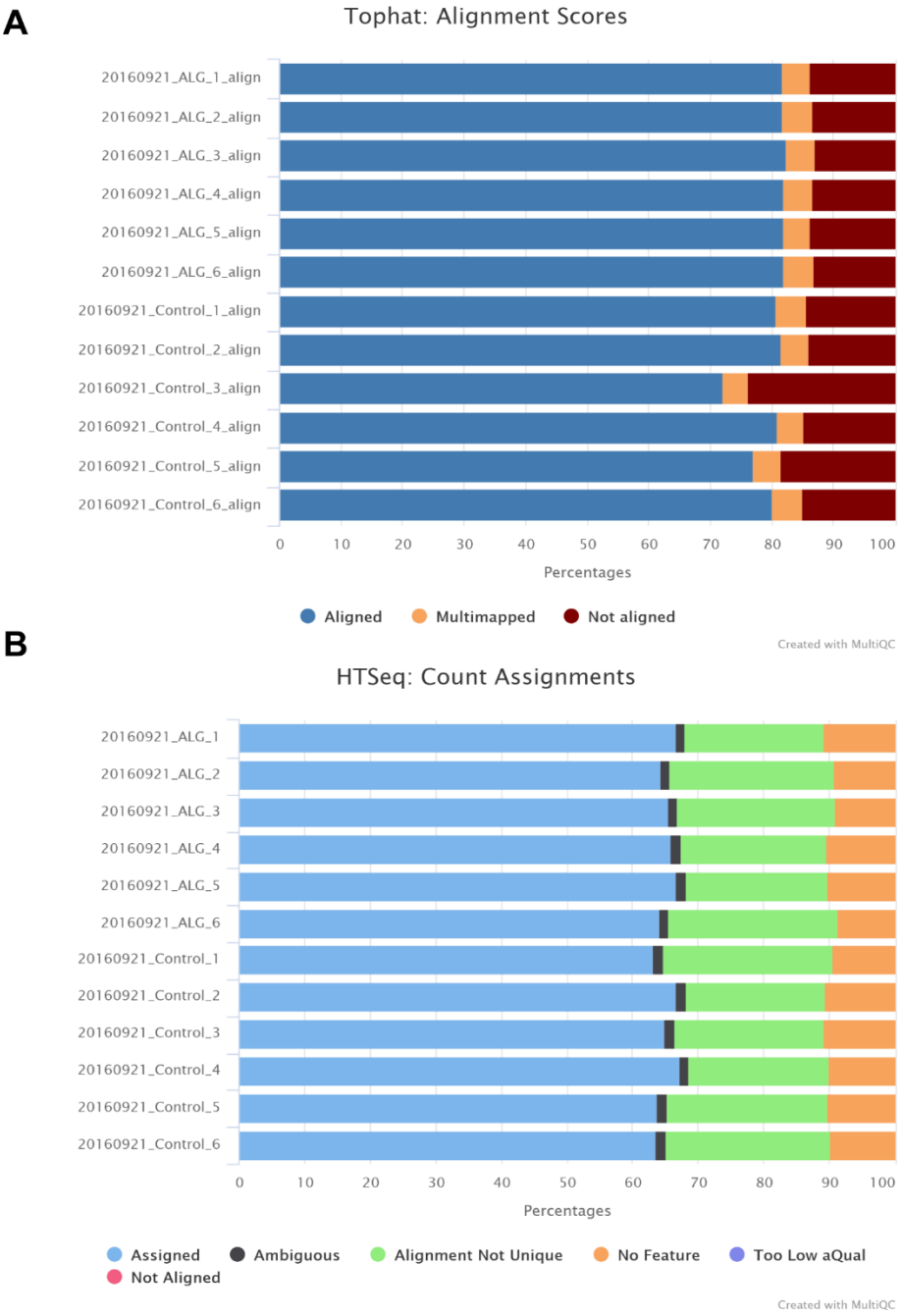

Figure 3.7 *Tophat2* and *HTSeq-count* statistics – PBMC samples
A. Between 76-87% of reads were mapped in all samples by *Tophat2*. **B.** *HTSeq-count* uniquely assigned (blue bar) between 63-67% of a sample's aligned reads to a gene. Trouble uniquely assigning reads likely due to difficulty with polymorphic antibody and T cell receptor genes.

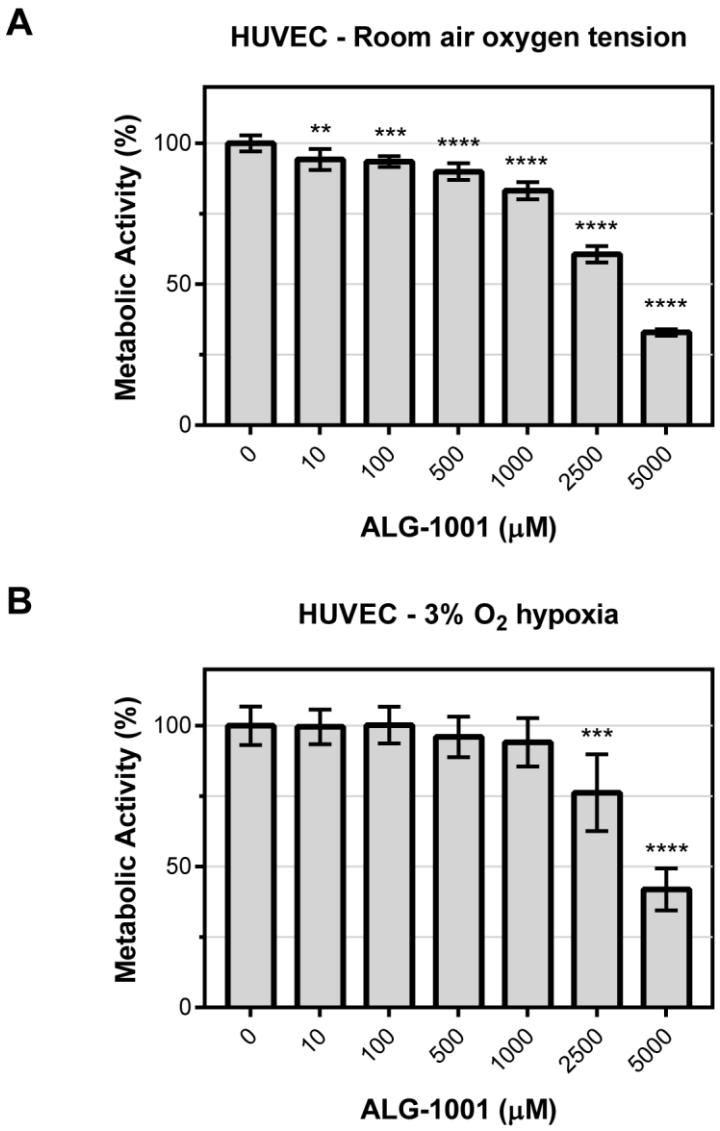

Figure 3.8 WST-1 assay of HUVECs treated with ALG-1001 concentration series
After 48 hours of ALG-1001 treatment at specified concentration, cellular metabolic activity level was measured as an indicator of cell density. At both room air oxygen tension (**A**) and 24 hours of 3% O2 hypoxia exposure (**B**) conditions, HUVECs treated with ALG-1001 demonstrated a dose-dependent drop in cell density at end of treatment. Experiment was performed once with n=6 per condition. Metabolic activity was normalized with mean activity of control cells as 100%. Mean and standard deviation plotted. All statistical tests were performed against control condition with the stars indicate the magnitude of the adjusted p-value (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$).

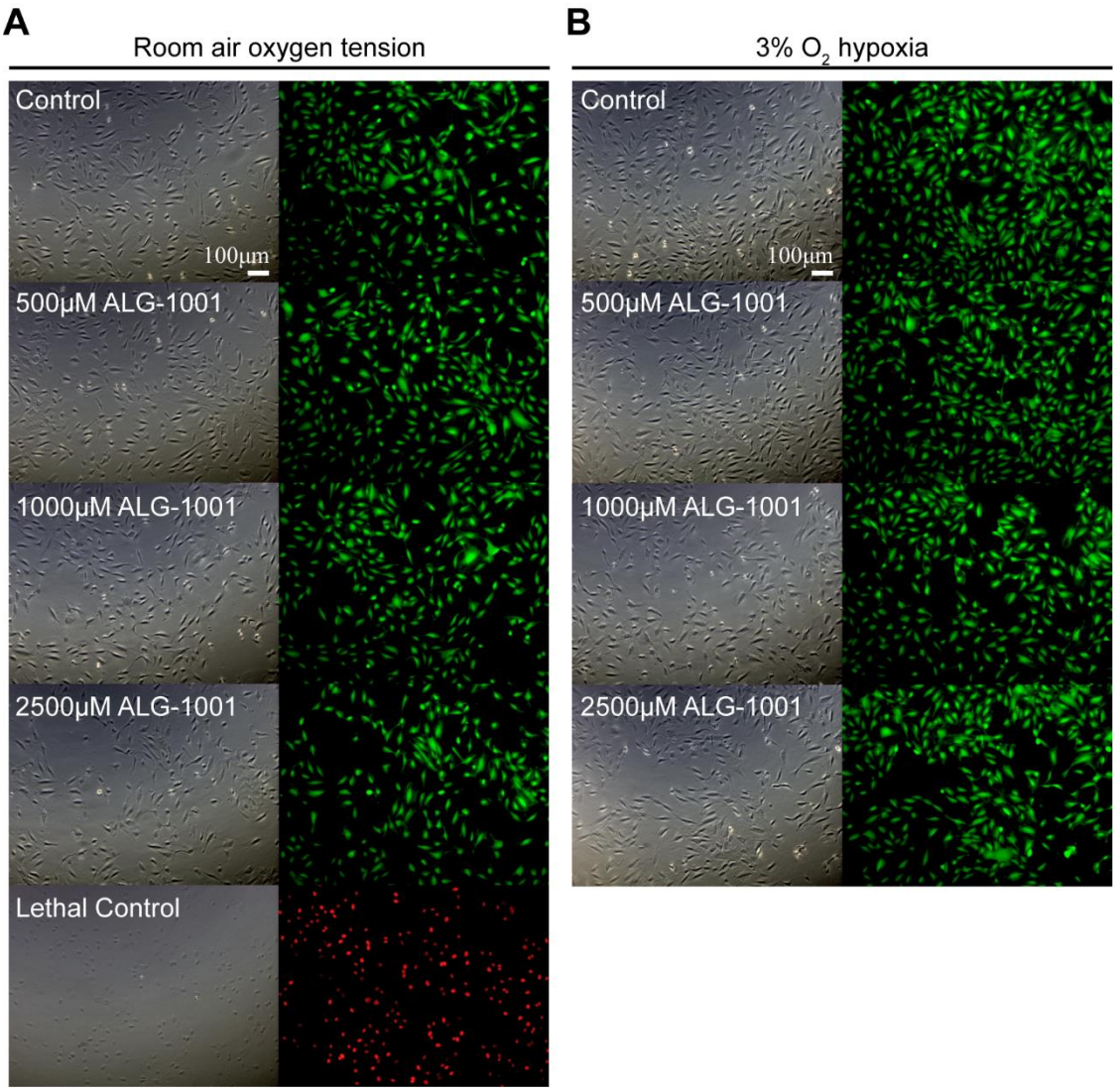

Figure 3.9 Dead and live staining of HUVECs treated with ALG-1001 concentration series

After 48 hours of ALG-1001 treatment at specified concentrations, dead and live staining was performed to detect cell state. (*left*) Phase contrast micrographs and (*right*) merged live (green) and dead (red) stained images show no obvious cell death or change in cell morphology after drug exposure. Experiment was repeated at both room air oxygen tension (**A**) and after 24 hours of 3% O_2 hypoxia (**B**) conditions. Lethal control shows cells that are dead with compromised cell membrane.

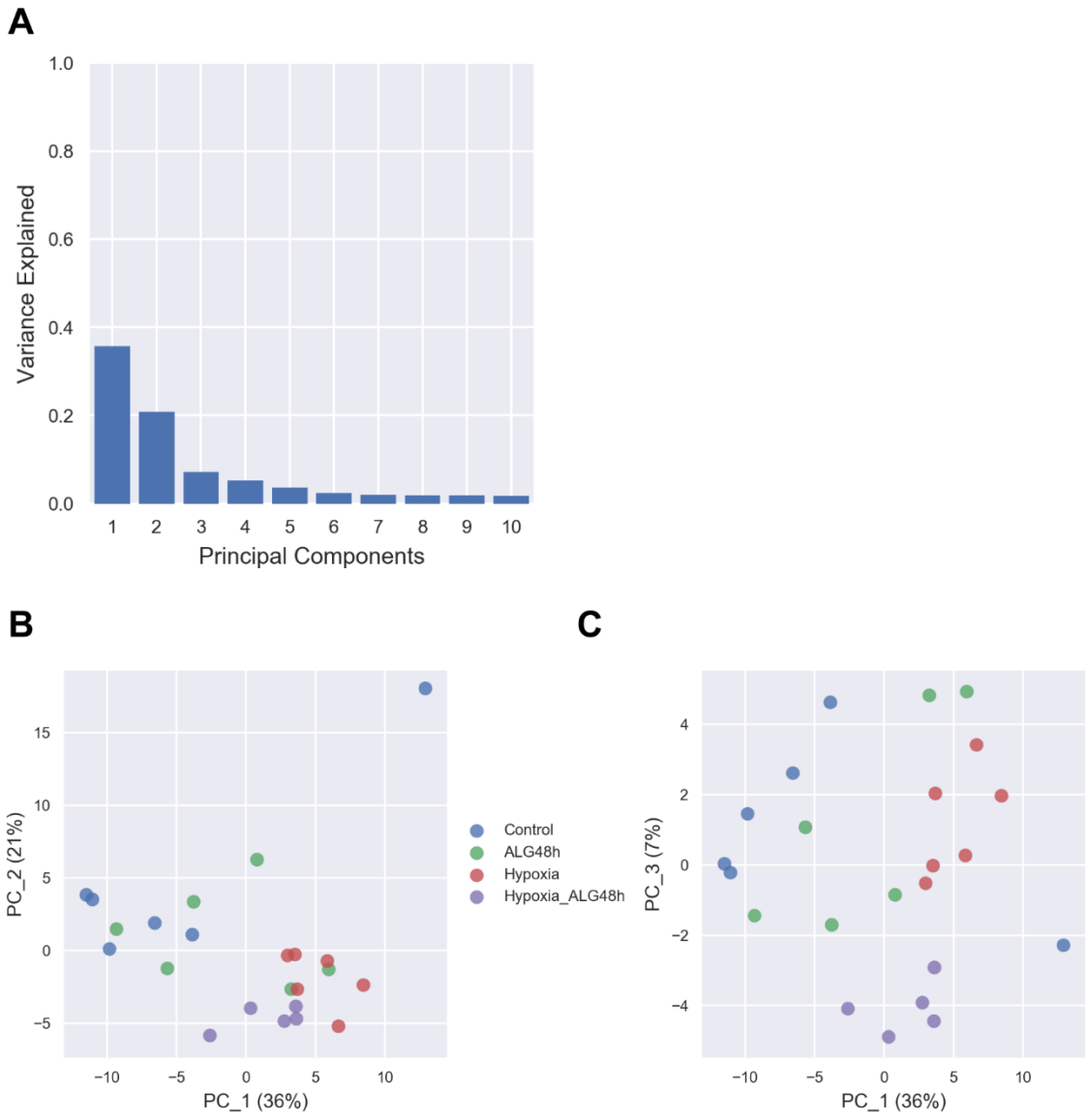

Figure 3.10 PCA of HUVEC, batch 1 samples
PCA of 24 HUVEC, batch 1 samples. ALG48h is ALG-1001, Hypoxia_ALG48h is ALG-1001+hypoxia conditions. PC_# is principal component #. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC_1 and PC_2, PC_1 and PC_3, respectively. Percent of variance explained are shown in axis label. Three obvious outliers were detected: control replicate 3, and ALG-1001 replicate 5 and 6.

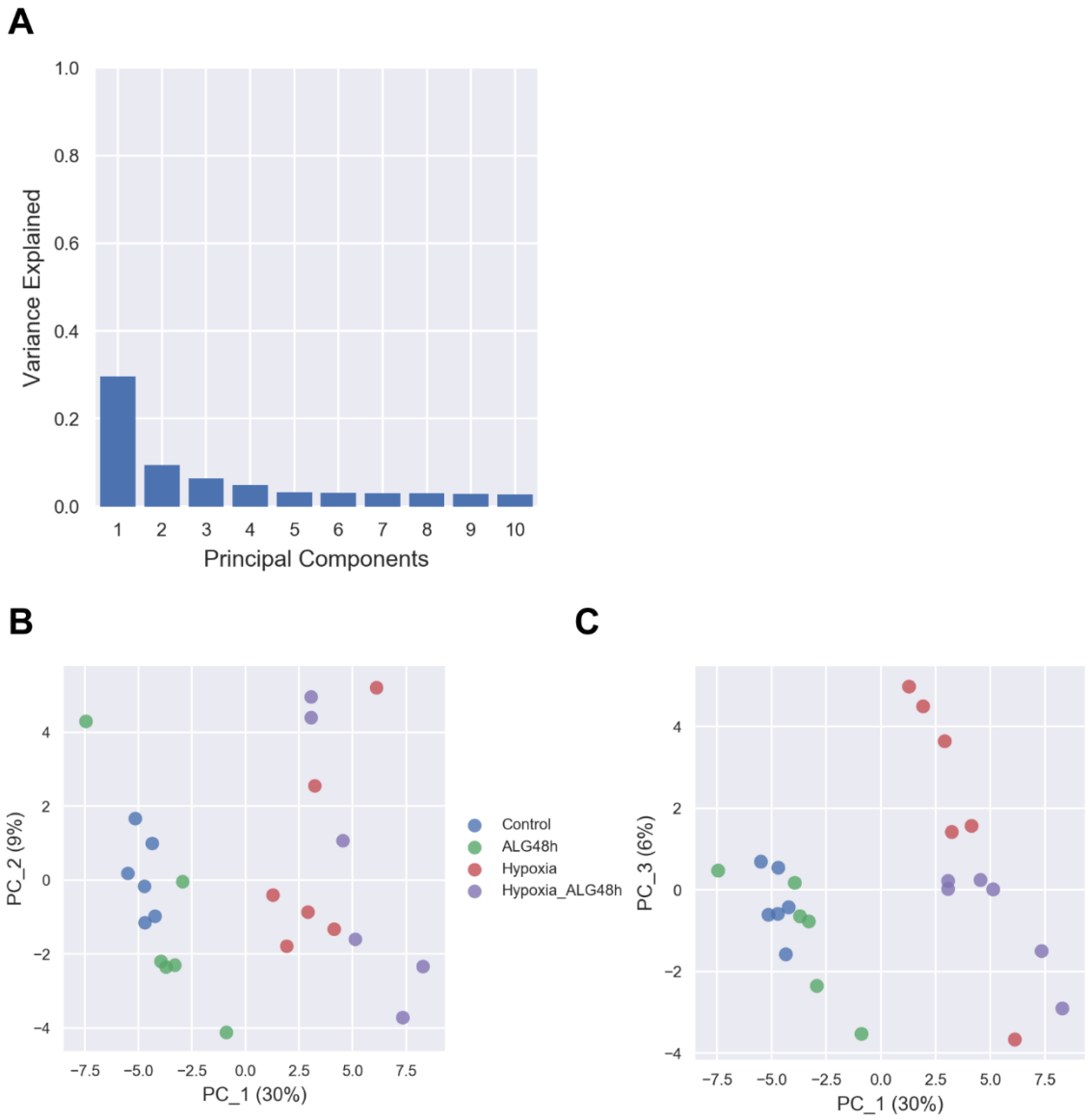

Figure 3.11 PCA of HUVEC, batch 2 samples
PCA of 24 HUVEC, batch 2 samples. ALG48h is ALG-1001, Hypoxia_ALG48h is ALG-1001+hypoxia conditions. PC_# is principal component #. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC_1 and PC_2, PC_1 and PC_3, respectively. Percent of variance explained are shown in axis label. Two obvious outliers were detected: ALG-1001 replicate 6 and hypoxia replicate 6.

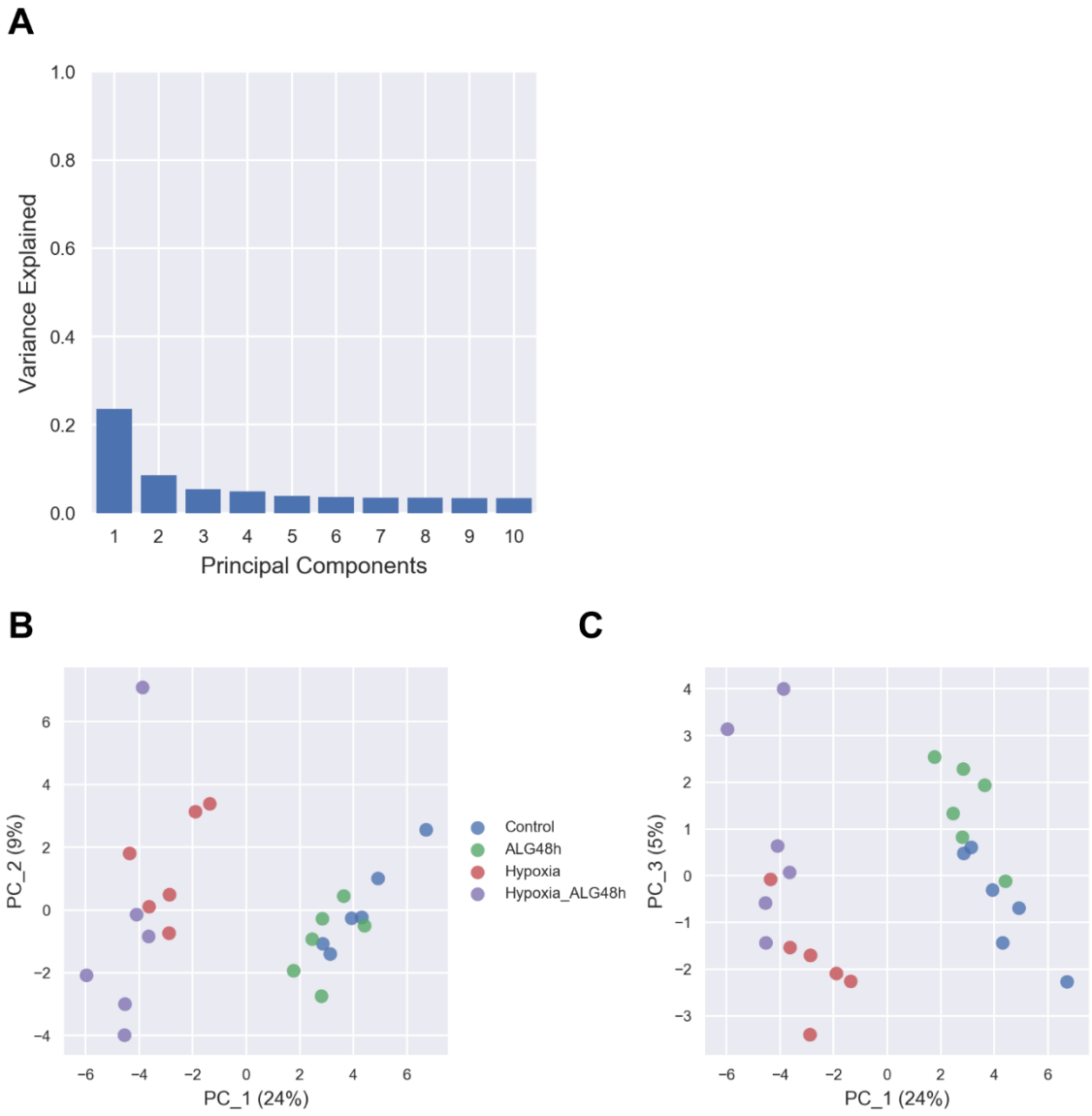

Figure 3.12 PCA of HUVEC, batch 3 samples
PCA of 24 HUVEC, batch 3 samples. ALG48h is ALG-1001, Hypoxia_ALG48h is ALG-1001+hypoxia conditions. PC_# is principal component #. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC_1 and PC_2, PC_1 and PC_3, respectively. Percent of variance explained are shown in axis label. One obvious outlier was detected: ALG-1001+hypoxia replicate 6.

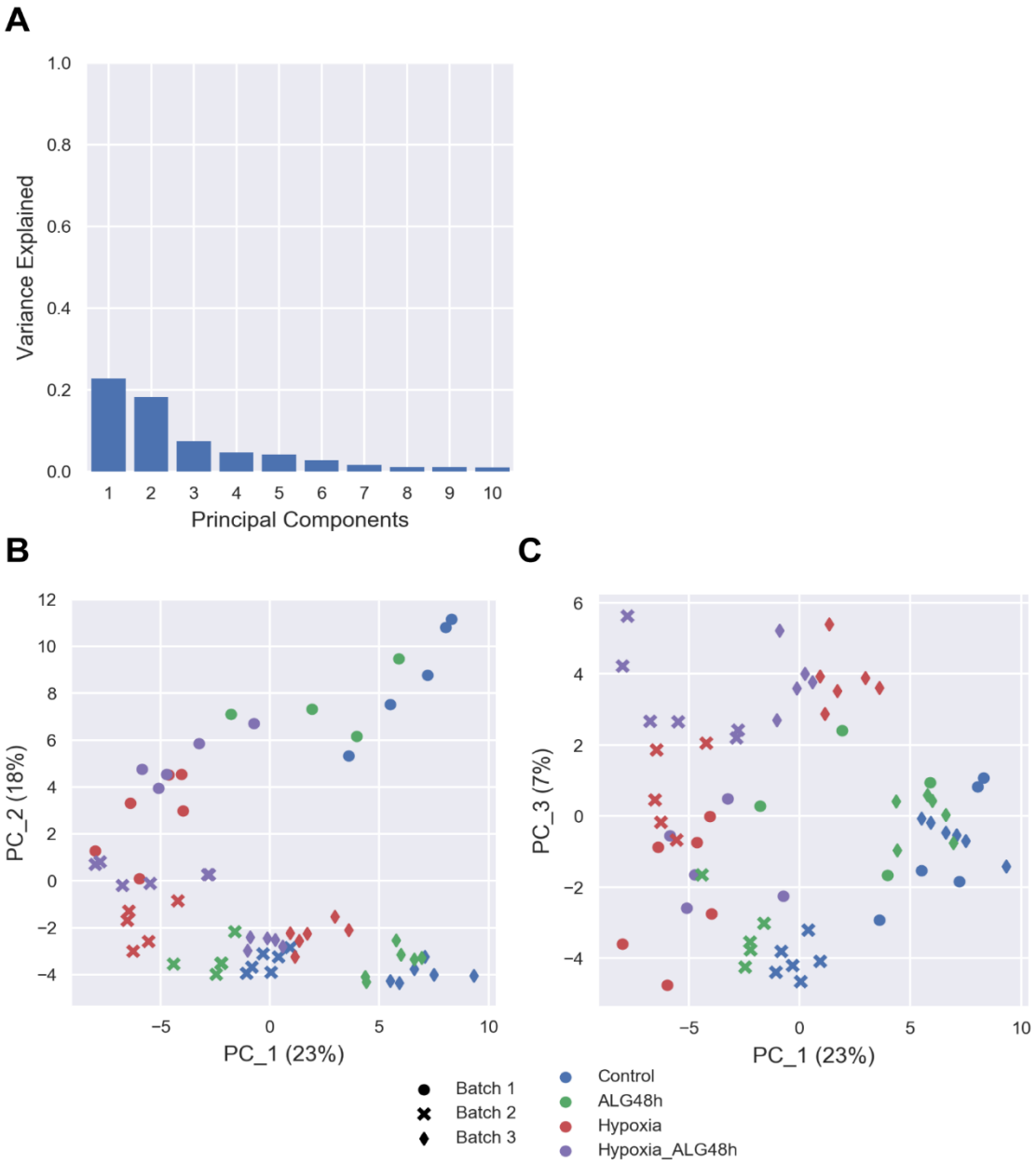

Figure 3.13 PCA of all non-outlier HUVEC samples
PCA of 66 non-outlier HUVEC samples. ALG48h is ALG-1001, Hypoxia_ALG48h is ALG-1001+hypoxia conditions. PC_# is principal component #. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC_1 and PC_2, PC_1 and PC_3, respectively. Percent of variance explained are shown in axis label. Experimental batch is symbol-coded while condition is color-coded.

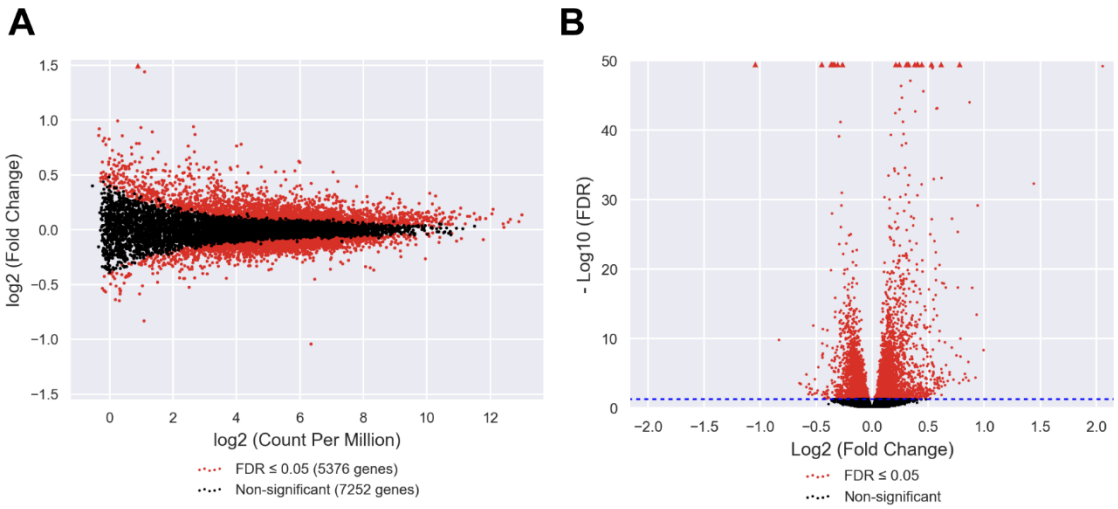

Figure 3.14 *edgeR* comparison of HUVEC control and hypoxia conditions
A. Measured CPM and fold-change for each expressed gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. **B**. *edgeR* measured fold-change and FDR displayed for each gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. Blue dashed line indicates FDR=0.05.

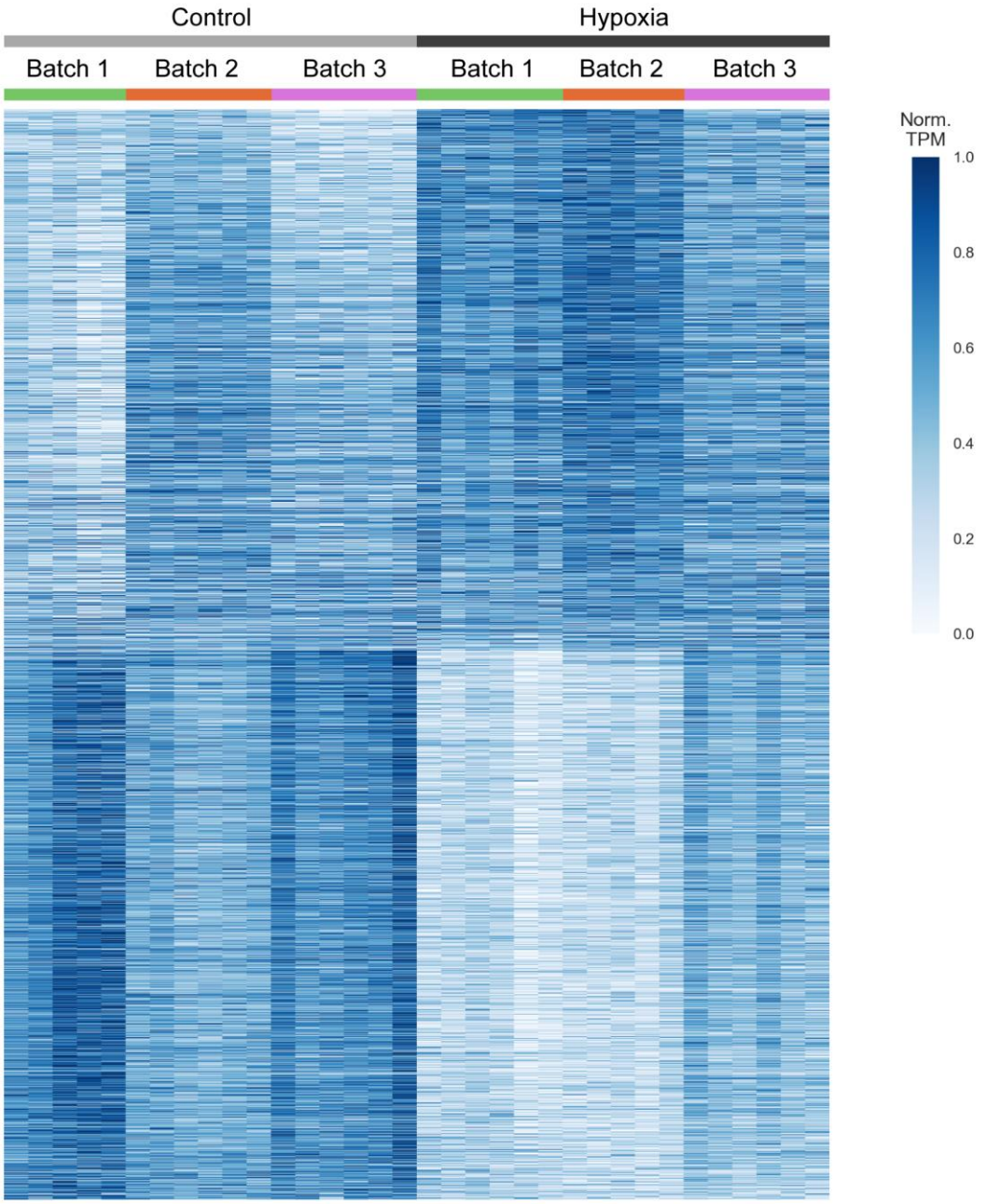

Figure 3.15 Heatmap of 5376 DE genes found to be regulated by hypoxia in HUVEC
Normalized (Norm.) TPM values are plotted for DE genes between HUVEC control and hypoxia conditions. Genes are plotted on rows and are ranked by decreasing log2-fold-change. Samples are grouped by condition (gray-scale) and by experimental batch (color) as indicated by the bars shown below the heatmap.

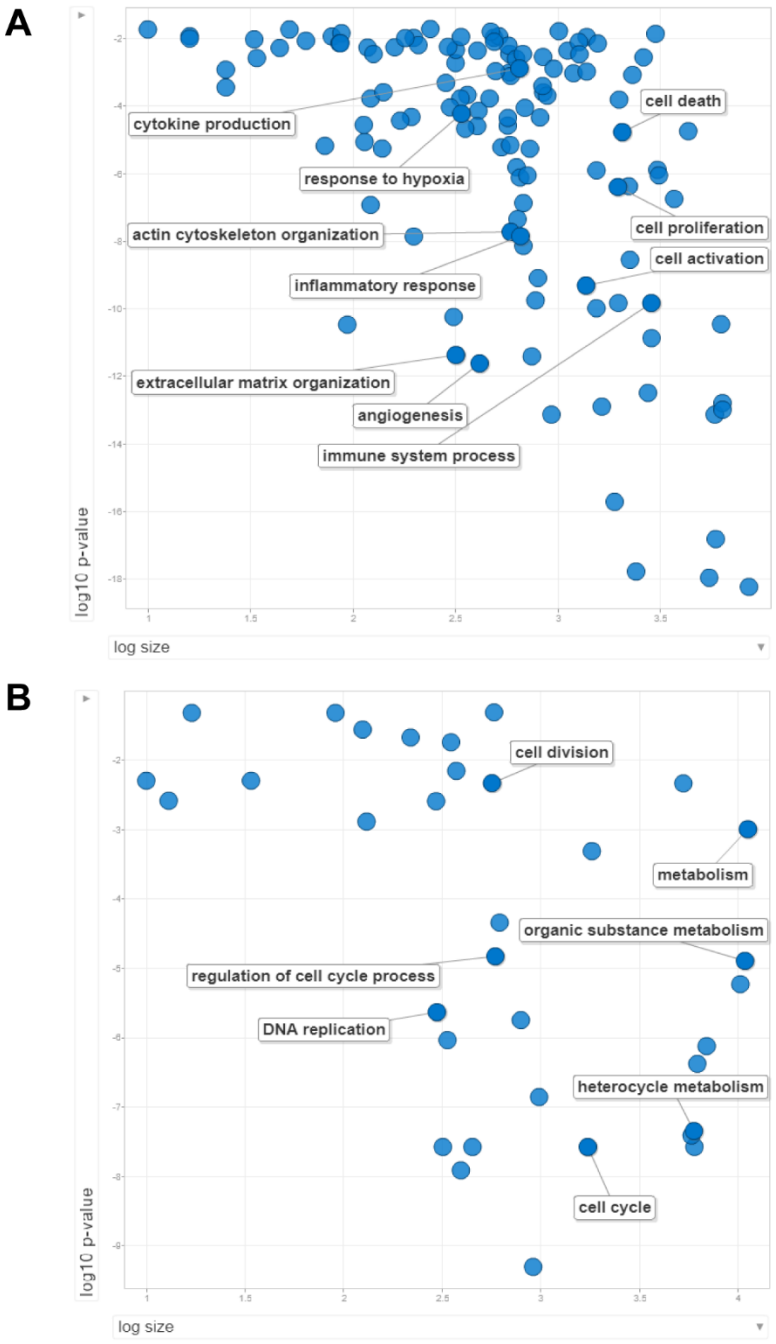

Figure 3.16 GO enrichment of DE genes regulated by hypoxia in HUVEC
Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO term. Selected biologically relevant GO terms were labeled.

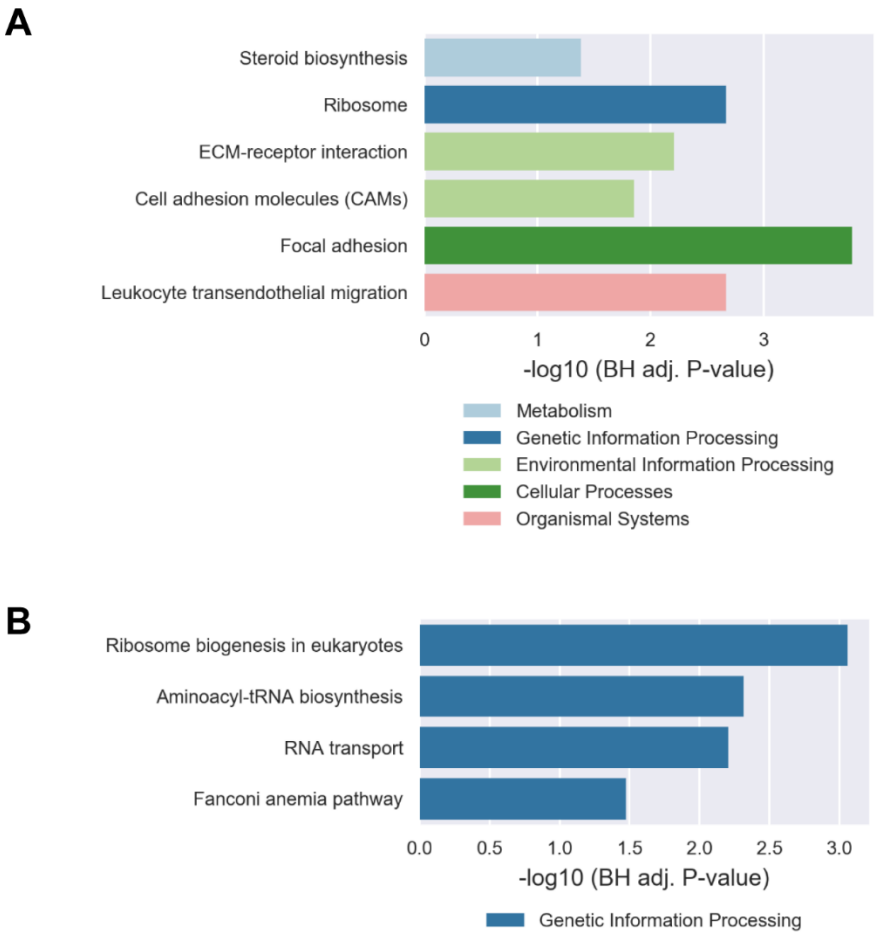

Figure 3.17 KEGG pathway enrichment of DE genes regulated by hypoxia in HUVEC *Goseq*-enriched KEGG pathways with up-regulated (**A**) or down-regulated (**B**) DE genes. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis.

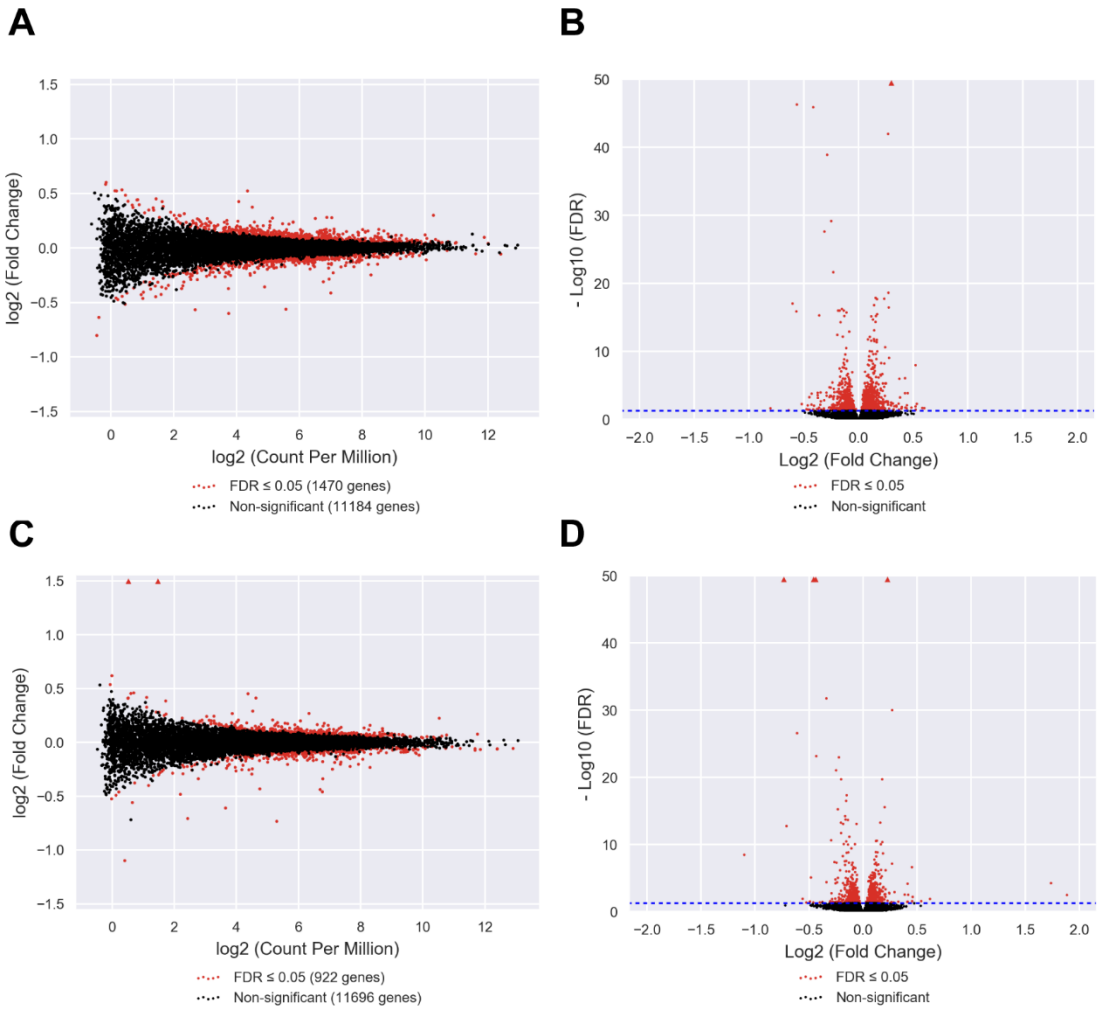

Figure 3.18 DE genes regulated by ALG-1001 at room air oxygen tension and hypoxia conditions

EdgeR used to identify DE genes between control and ALG-1001 conditions (**A**-**B**) and hypoxia and ALG-1001 + hypoxia conditions (**C**-**D**). (**A**, **C**) Measured CPM and fold-change for each expressed gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. (**B**, **D**) E*dgeR* measured fold-change and FDR displayed for each gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. Blue dashed line indicates FDR=0.05.

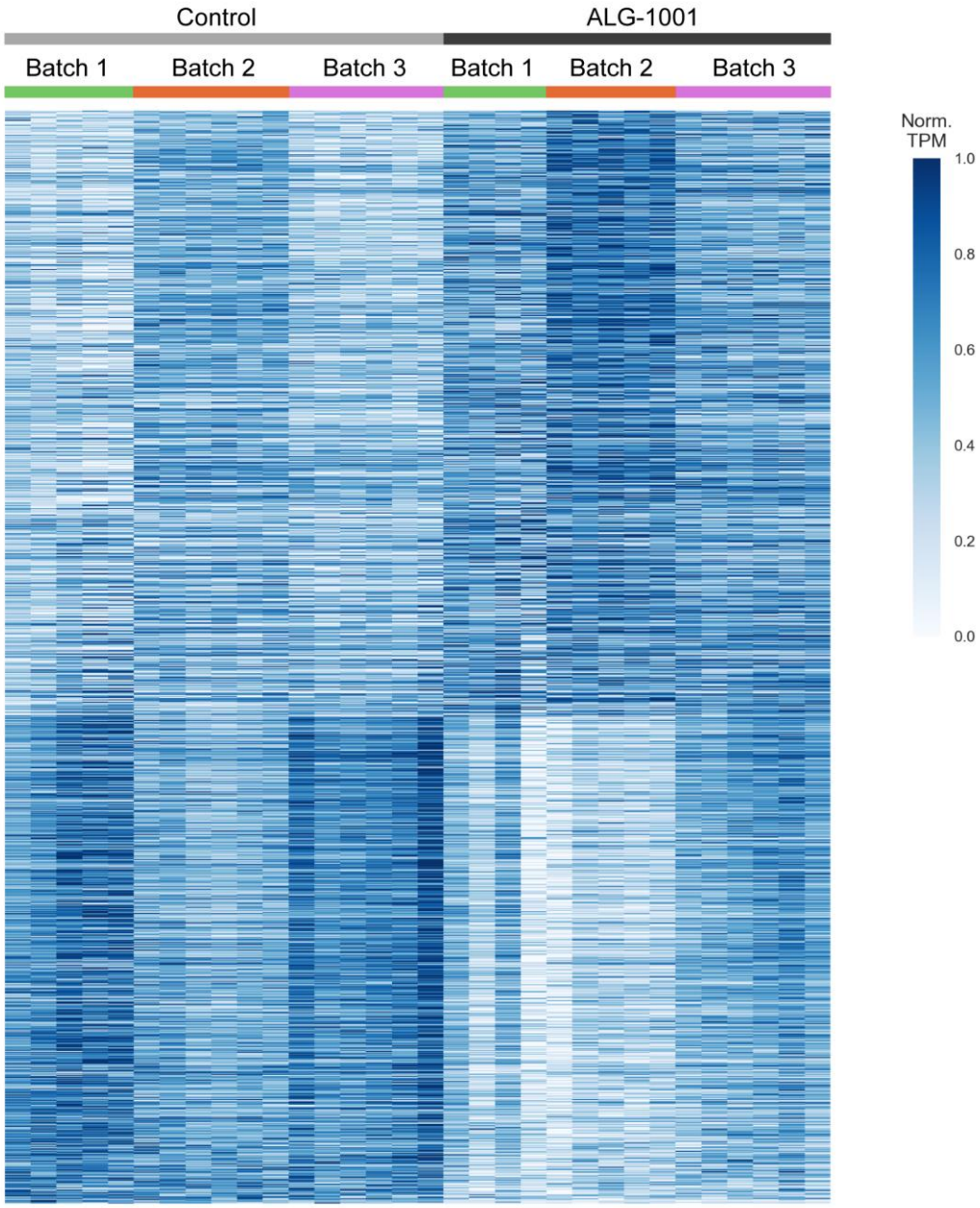

Figure 3.19 Heatmap of 1470 DE genes regulated by ALG-1001 at room air oxygen tension in HUVEC

Normalized (Norm.) TPM values are plotted for DE genes between HUVEC control and ALG-1001 conditions. Genes are plotted on rows and are ranked by decreasing log2-fold-change. Samples are grouped by condition (gray-scale) and by experimental batch (color) as indicated by the bars shown below the heatmap.

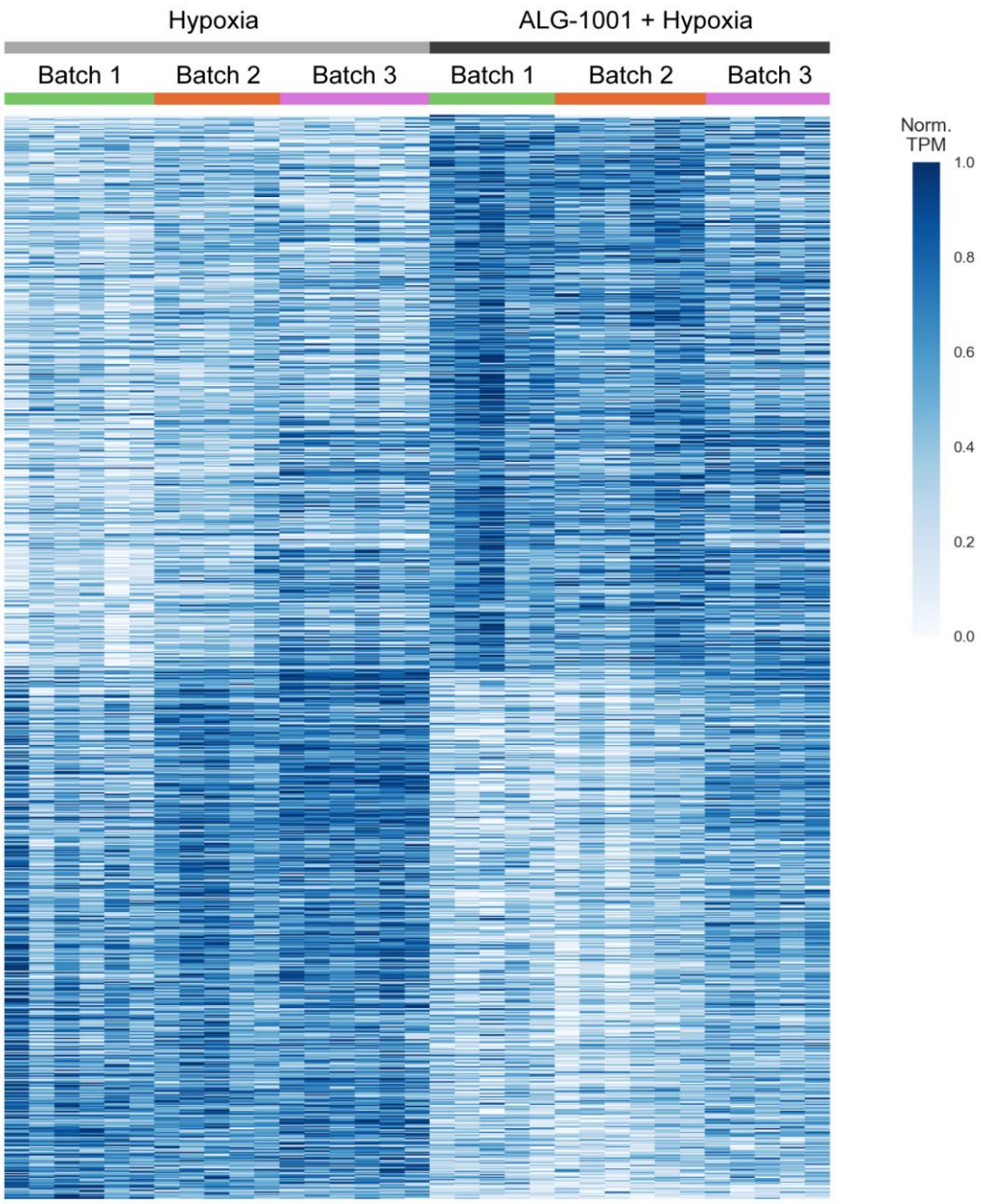

Figure 3.20 Heatmap of 922 DE genes regulated by ALG-1001 with hypoxia co-treatment in HUVEC

Normalized (Norm.) TPM values are plotted for DE genes between HUVEC hypoxia and ALG-1001 + hypoxia conditions. Genes are plotted on rows and are ranked by decreasing log2-fold-change. Samples are grouped by condition (gray-scale) and by experimental batch (color) as indicated by the bars shown below the heatmap.

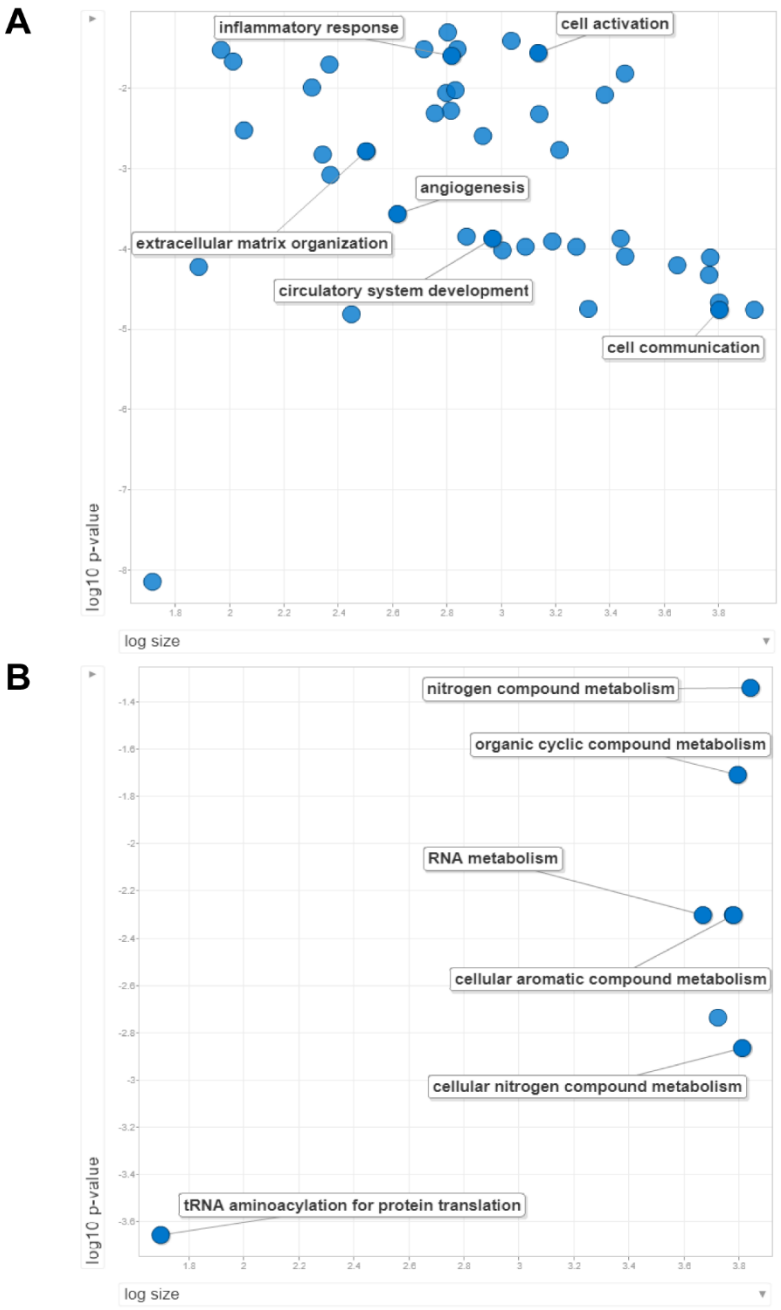

Figure 3.21 GO enrichment of DE genes regulated by ALG-1001 at room air oxygen tension

Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO term. Selected biologically relevant GO terms were labeled.

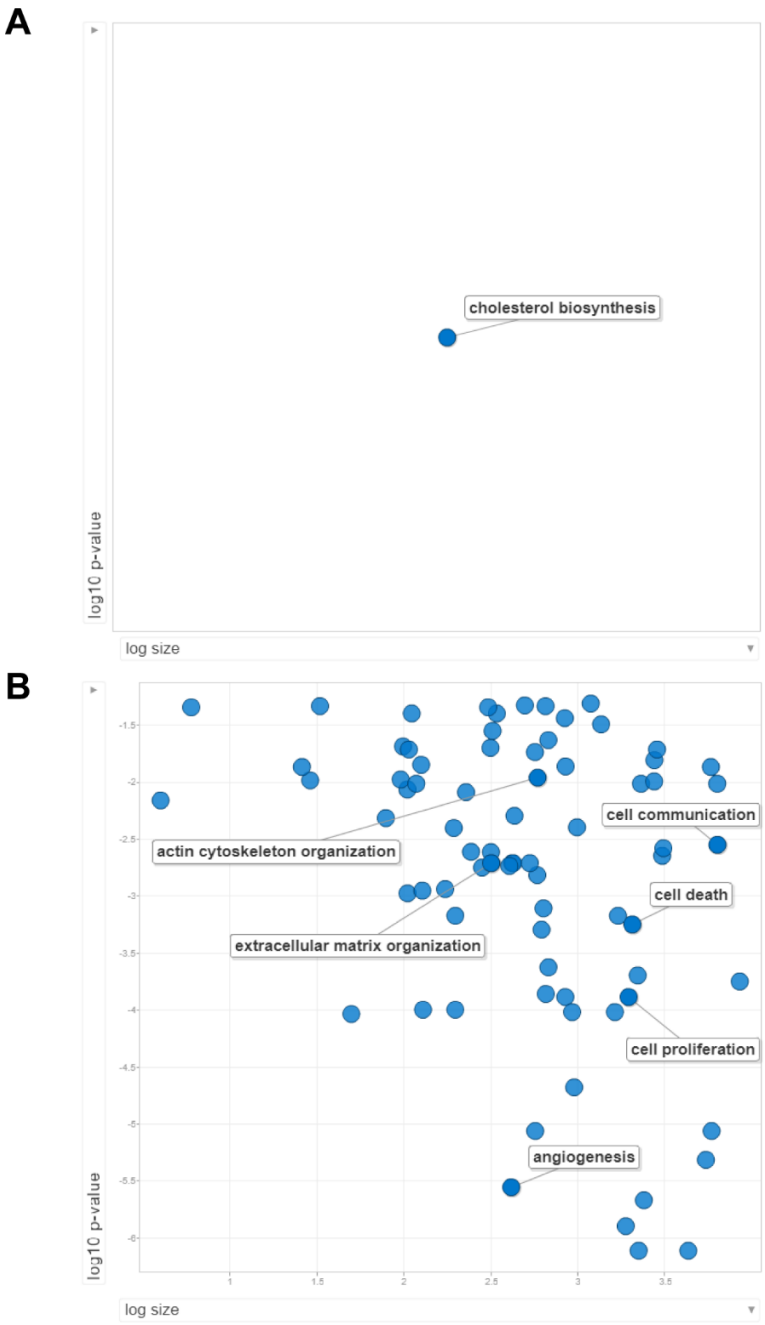

Figure 3.22 GO enrichment of DE genes regulated by ALG-1001 with hypoxia stimulation

Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO term. Selected biologically relevant GO terms were labeled.

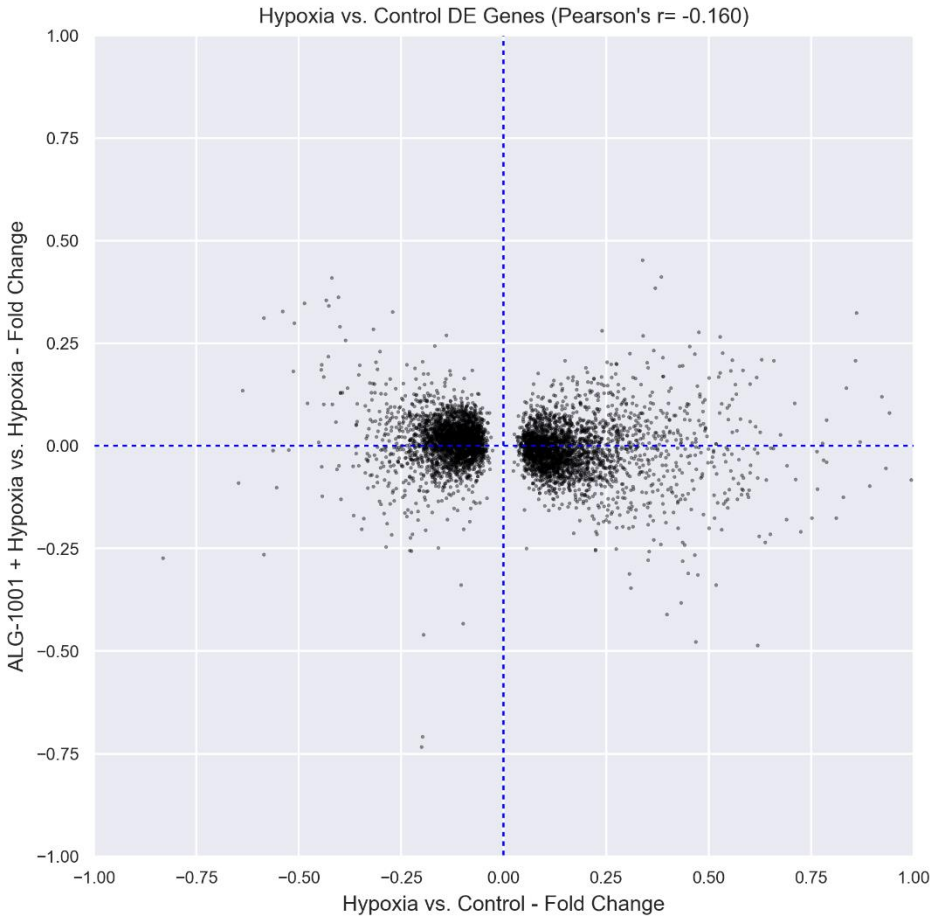

Figure 3.23 ALG-1001 treatment weakly moderates hypoxia-induced transcriptome regulation

Each dot represents a DE gene found to be regulated during hypoxia response. The gene's log2-fold-change after hypoxia treatment is plotted on the x-axis, while the fold change with ALG-1001 co-treatment is plotted on the y-axis. Plot window is limited to the (-1, 1) range on both axes. Pearson's r value is measured as -0.160. The plot can be divided into four quadrants, as indicated by the blue dashed lines, with 1710 genes in the top left, 986 in the bottom left, 1106 in the top right, and 1568 in the bottom right.

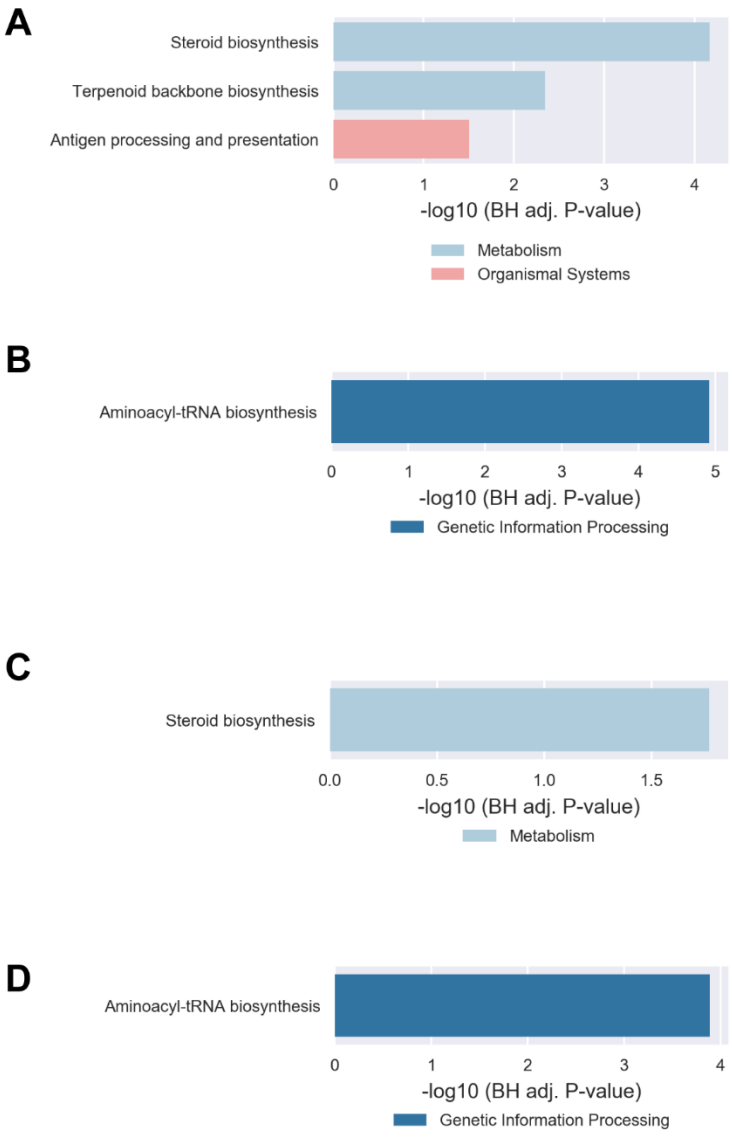

Figure 3.24 KEGG pathway enrichment of DE genes regulated by ALG-1001 at room air oxygen tension and hypoxia conditions

Goseq-enriched KEGG pathways with up-regulated (**A**) or down-regulated (**B**) DE genes by ALG-1001 at room air oxygen tension and up-regulated (**C**) or down-regulated (**D**) DE genes by ALG-1001 with hypoxia stimulation. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis.

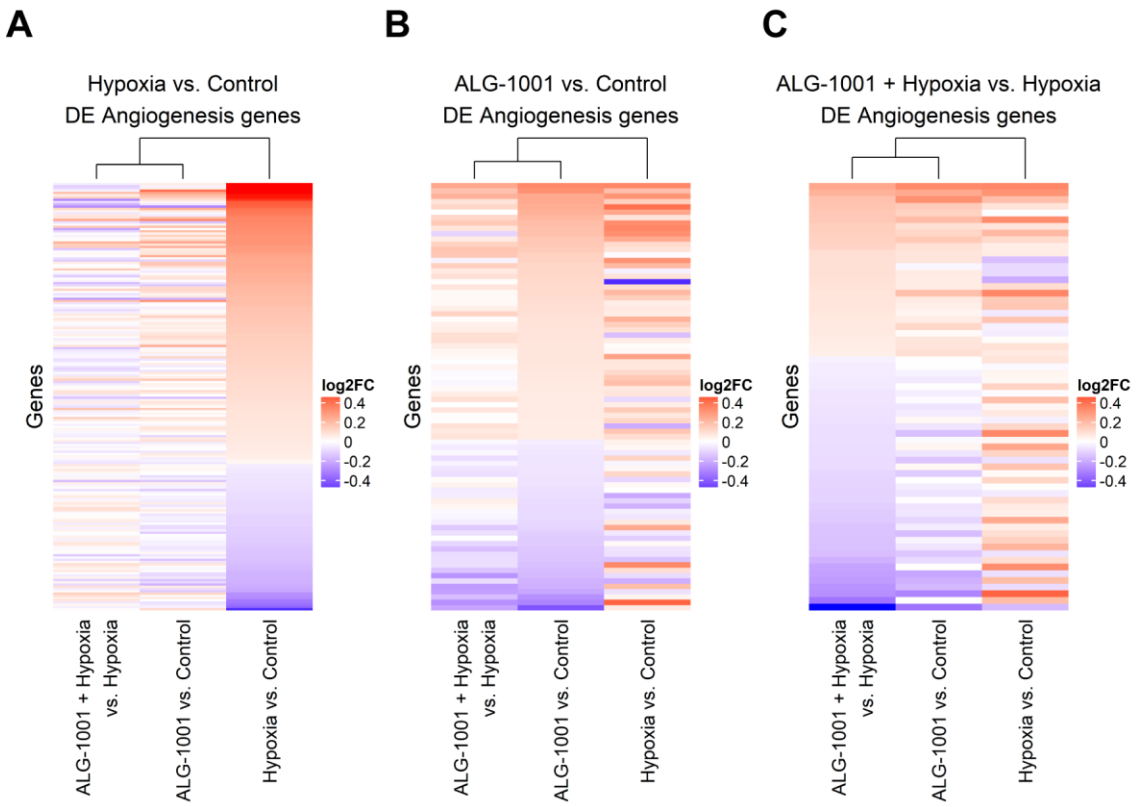

Figure 3.25 Heatmap of DE angiogenesis genes

Fold change of DE "angiogenesis" (GO biological process) genes are plotted as heatmap. Genes are plotted row-wise and comparisons column-wise. Expression profiles are clustered using average Euclidean distance method with results shown above the heatmap. |log2-fold-change|>0.50 are shown as 0.50 or -0.50 to improve visualization of small changes. Log2FC is log2-fold-change**.** DE "angiogenesis" genes found in hypoxia vs. control (**A**), ALG-1001 vs. control (**B**) and ALG-1001 + hypoxia vs. hypoxia (**C**) comparisons are plotted, with genes ranked by the log2-fold-change in each respective comparison.

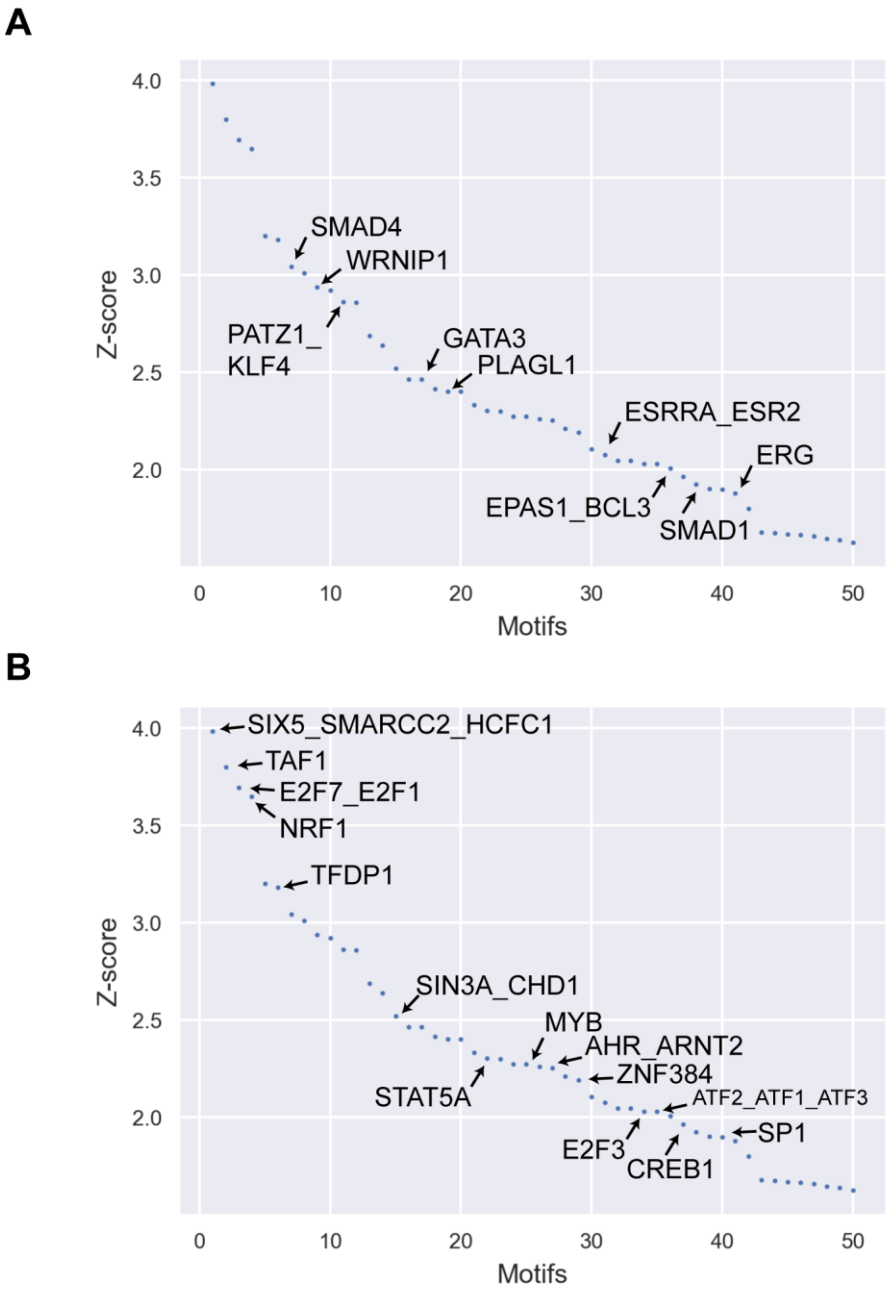

Figure 3.26 *ISMARA* prediction of TFs involved during hypoxia response
HUVEC control and hypoxia samples were submitted to *ISMARA* for enrichment of TFs that are likely involved during hypoxia-response. Top 50 enriched factors with the highest Z-score were plotted (x-axis shows rank by Z-score). Selected biologically relevant factors with higher (**A**) or lower (**B**) target gene activity in hypoxia condition are labeled.

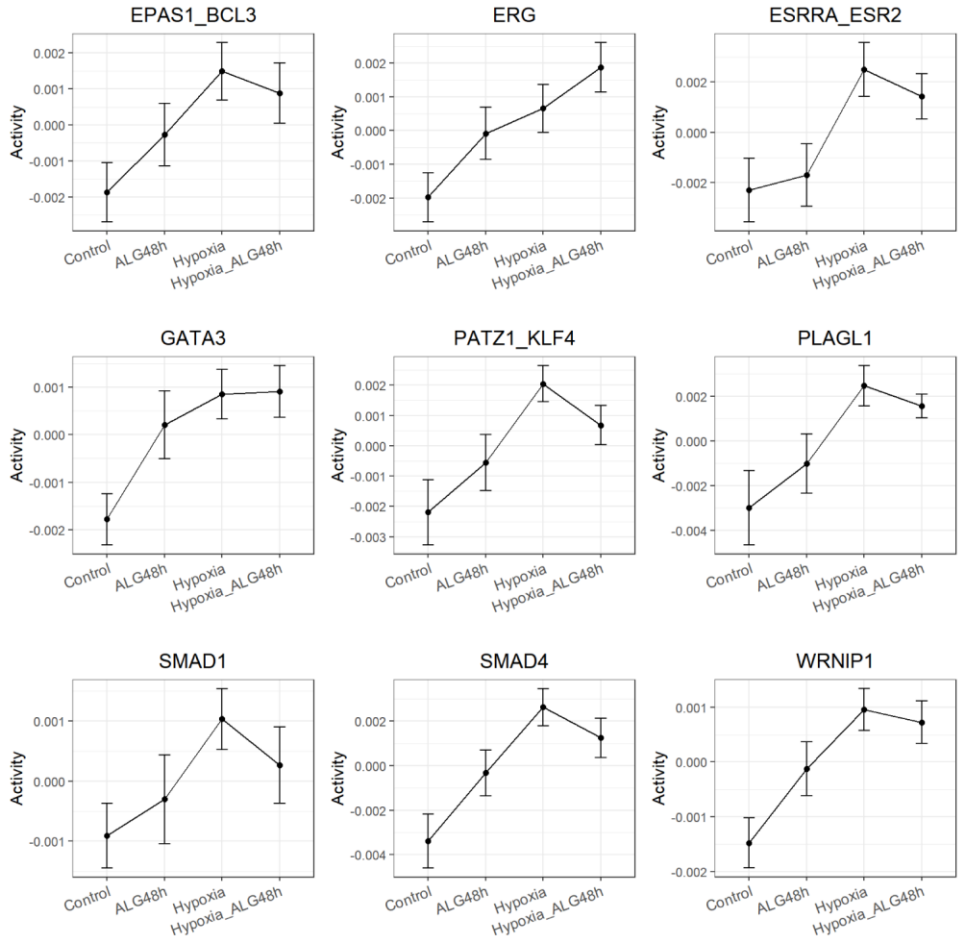

Figure 3.27 Activity profile of selected TF motifs with higher target activity in hypoxia samples

HUVEC control and hypoxia samples were submitted to *ISMARA* for enrichment of TFs that are likely involved during hypoxia-response. Activity profile (mean + standard deviation) of selected TFs with higher target activity in hypoxia condition are plotted. Conditions labeled as follows: ALG-1001 as ALG48h, ALG-1001 + hypoxia as Hypoxia_ALG48h.

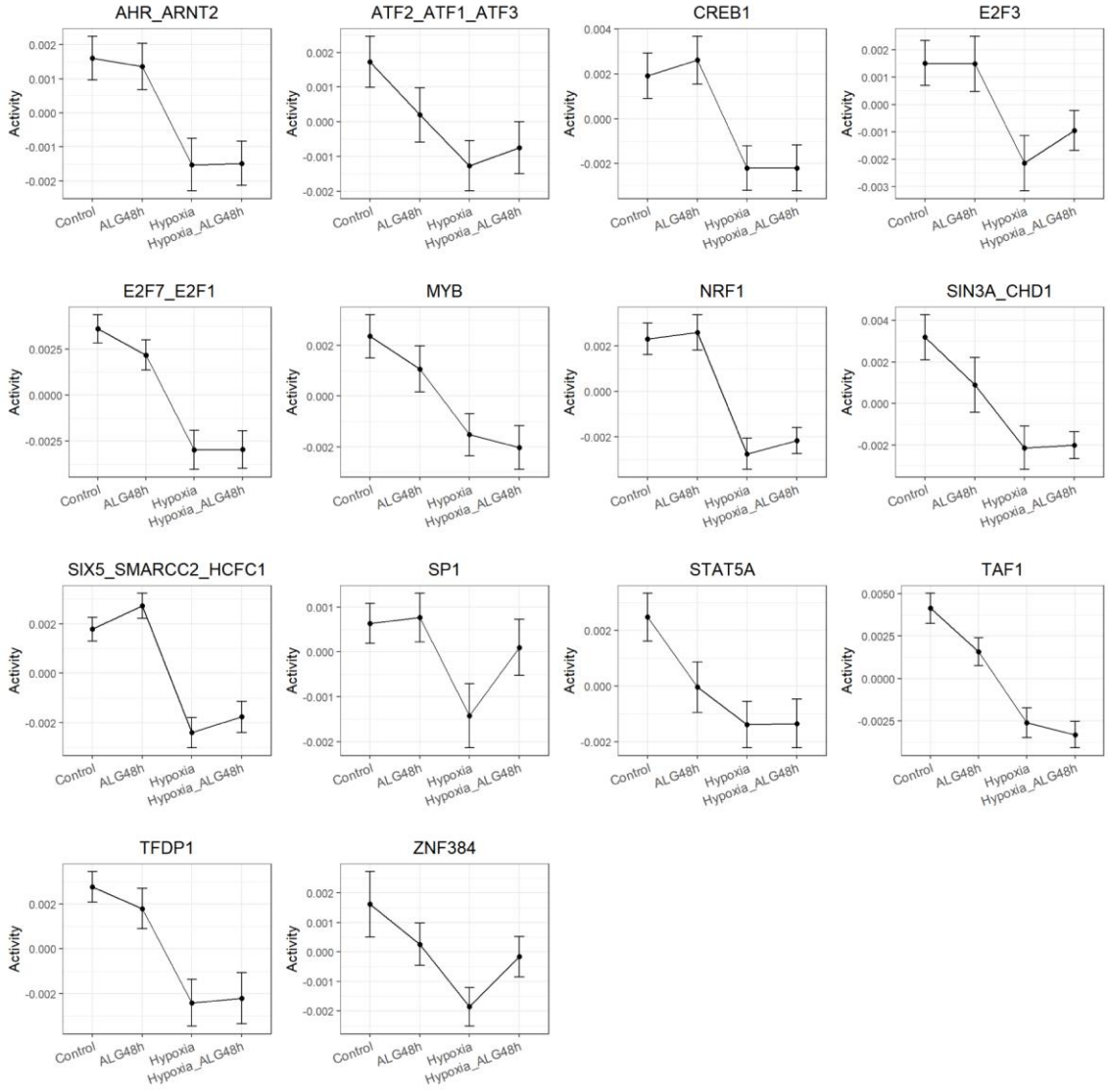

Figure 3.28 Activity profile of selected TF motifs with lower target activity in hypoxia samples

HUVEC control and hypoxia samples were submitted to *ISMARA* for enrichment of TFs that are likely involved during hypoxia-response. Activity profile (mean + standard deviation) of selected TFs with lower target activity in hypoxia condition are plotted. Conditions labeled as follows: ALG-1001 as ALG48h, ALG-1001 + hypoxia as Hypoxia_ALG48h.

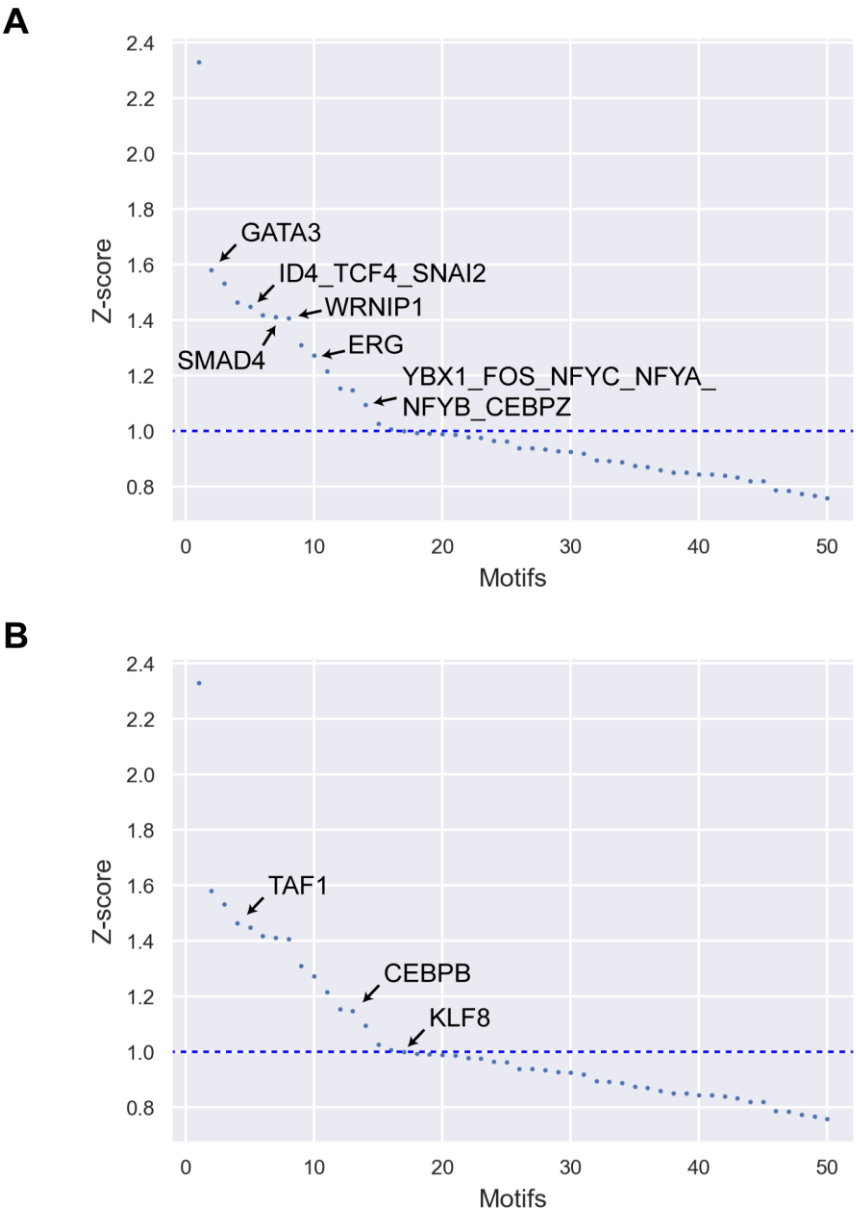

Figure 3.29 *ISMARA* prediction of TFs regulated by ALG-1001 at room air oxygen tension

HUVEC control and ALG-1001 samples were submitted to *ISMARA* for enrichment of TFs that are likely involved during ALG-1001 induced transcriptome regulation. Top 50 enriched factors with the highest Z-score were plotted (x-axis shows rank by Z-score). Only results with Z-score > 1.0 (blue dashed line) are considered for further analysis. Selected biologically relevant factors with higher (**A**) or lower (**B**) target gene activity in drug treatment condition are labeled.

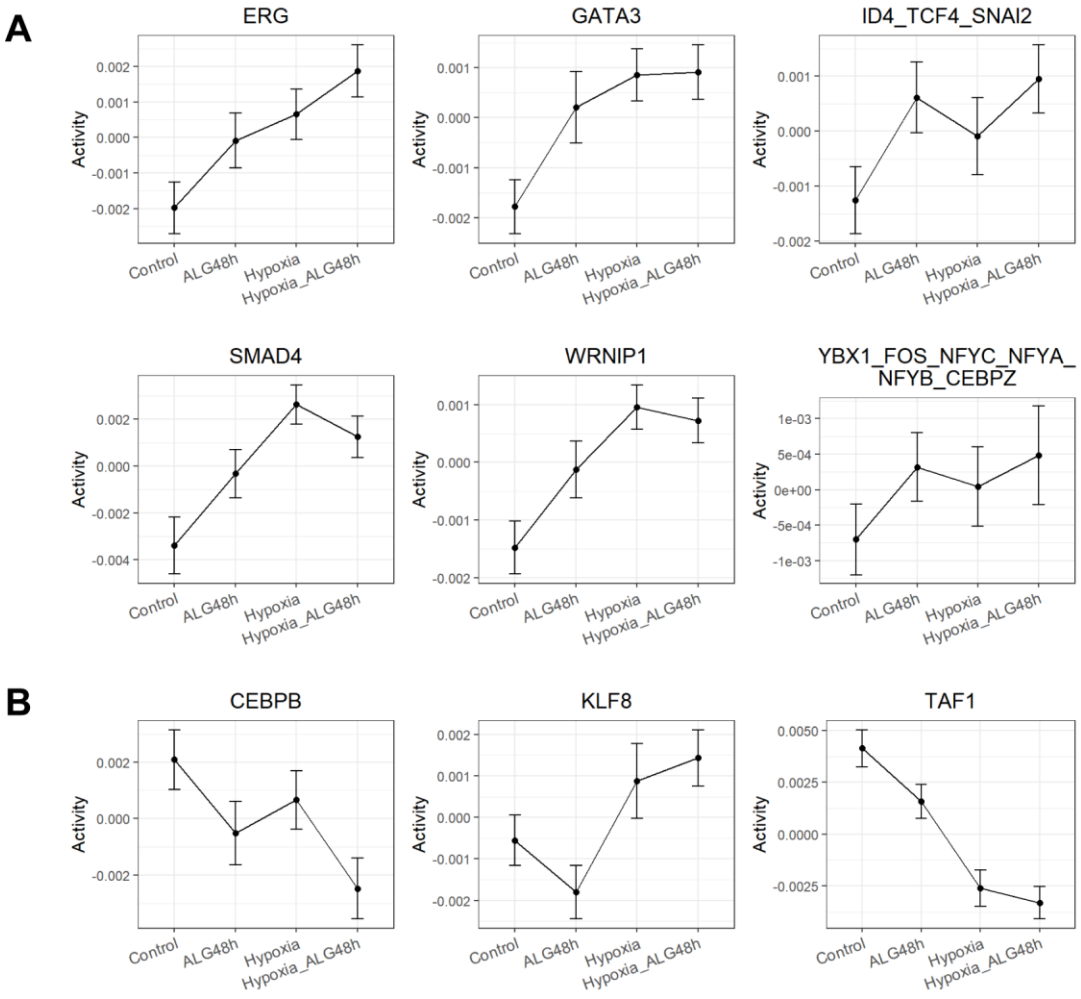

Figure 3.30 Activity profile of selected TF motifs likely involved in ALG-1001 induced transcriptome regulation

HUVEC control and ALG-1001 samples were submitted to *ISMARA* for enrichment of TFs that are likely involved during ALG-1001 response. Activity profile (mean + standard deviation) of selected TFs with higher (**A**) or lower (**B**) target activity in ALG-1001 condition are plotted. Conditions labeled as follows: ALG-1001 as ALG48h, ALG-1001 + hypoxia as Hypoxia_ALG48h.

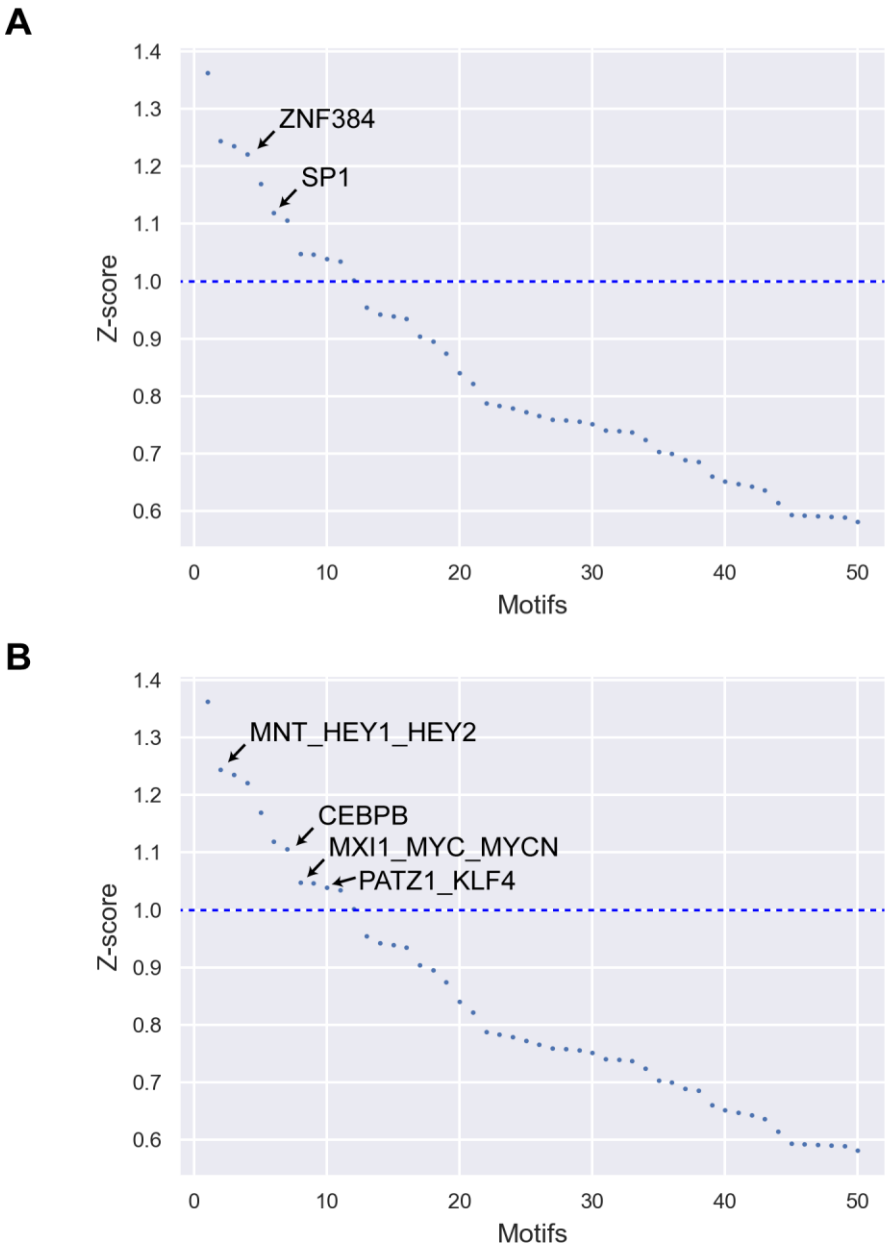

Figure 3.31 *ISMARA* prediction of TFs regulated by ALG-100 with hypoxia treatment
HUVEC hypoxia and ALG-1001 + hypoxia samples were submitted to *ISMARA* for enrichment of TFs that are likely involved during ALG-1001 induced transcriptome regulation. Top 50 enriched factors with the highest Z-score were plotted (x-axis shows rank by Z-score). Only results with Z-score > 1.0 (blue dashed line) are considered for further analysis. Selected biologically relevant factors with higher (**A**) or lower (**B**) target gene activity in drug treatment condition are labeled.

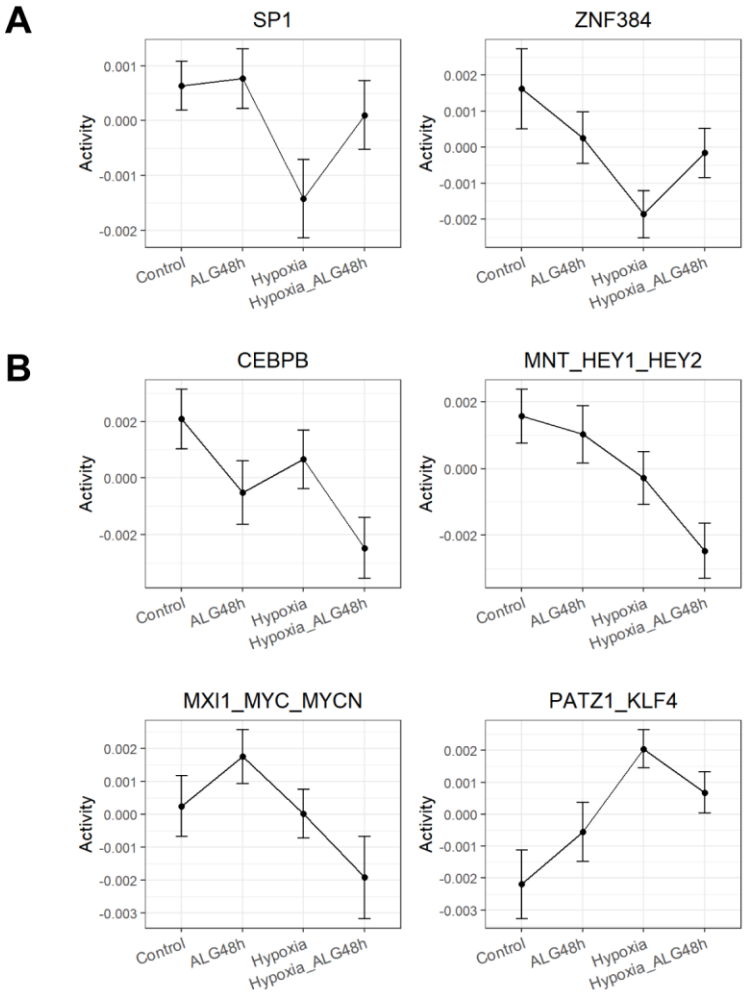

Figure 3.32 Activity profile of selected TF motifs likely regulated by ALG-1001 with hypoxia treatment

HUVEC hypoxia and ALG-1001 + hypoxia samples were submitted to *ISMARA* for enrichment of TFs that are likely involved during ALG-1001 response. Activity profile (mean + standard deviation) of selected TFs with higher (**A**) or lower (**B**) target activity in ALG-1001 + hypoxia condition are plotted. Conditions labeled as follows: ALG-1001 as ALG48h, ALG-1001 + hypoxia as Hypoxia_ALG48h.

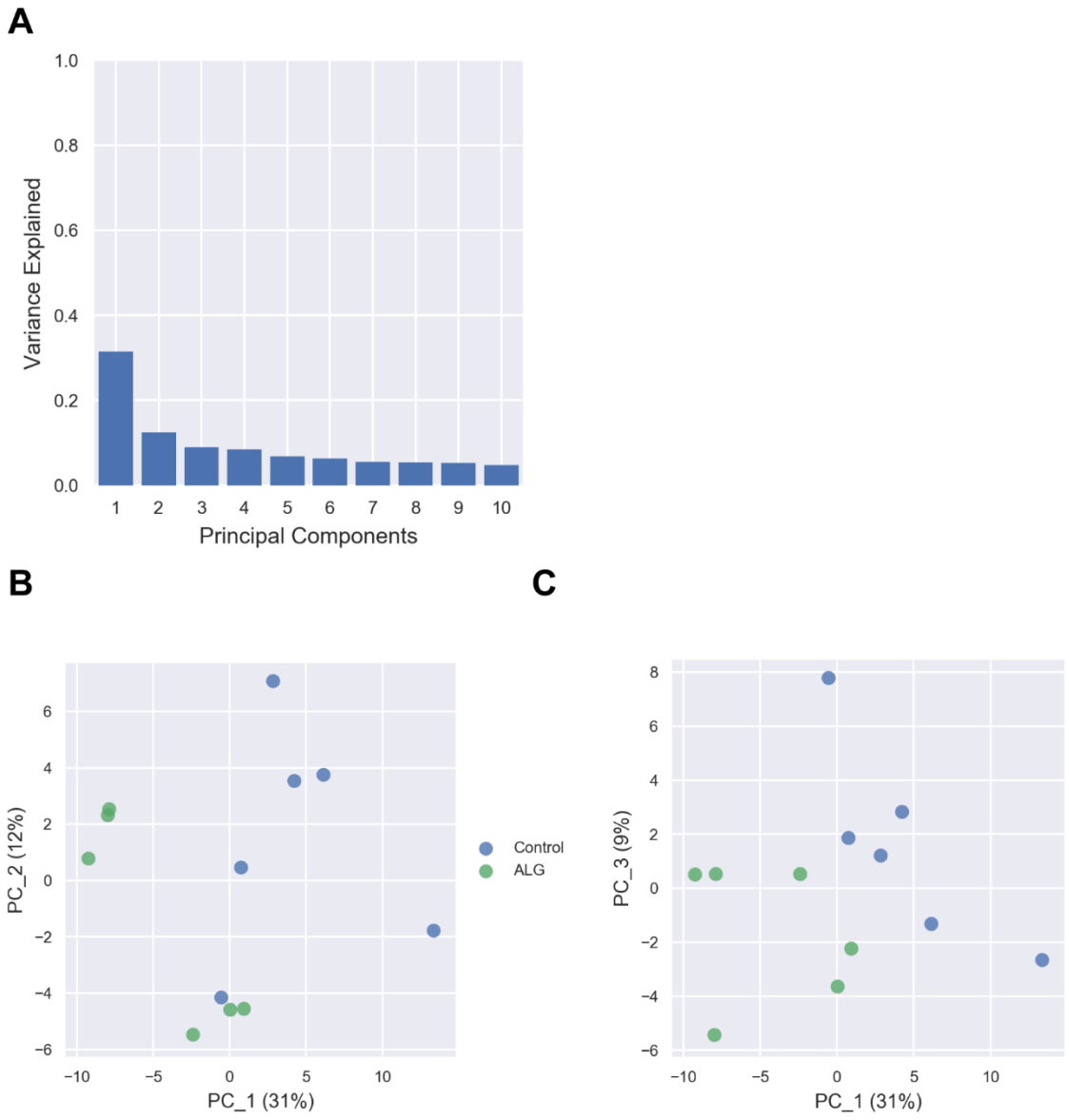

Figure 3.33 PCA of PBMC samples
PCA of 12 PBMC samples. ALG-1001 labeled as ALG. PC_# is principal component #. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC_1 and PC_2, PC_1 and PC_3, respectively. Percent of variance explained is shown in axis label.

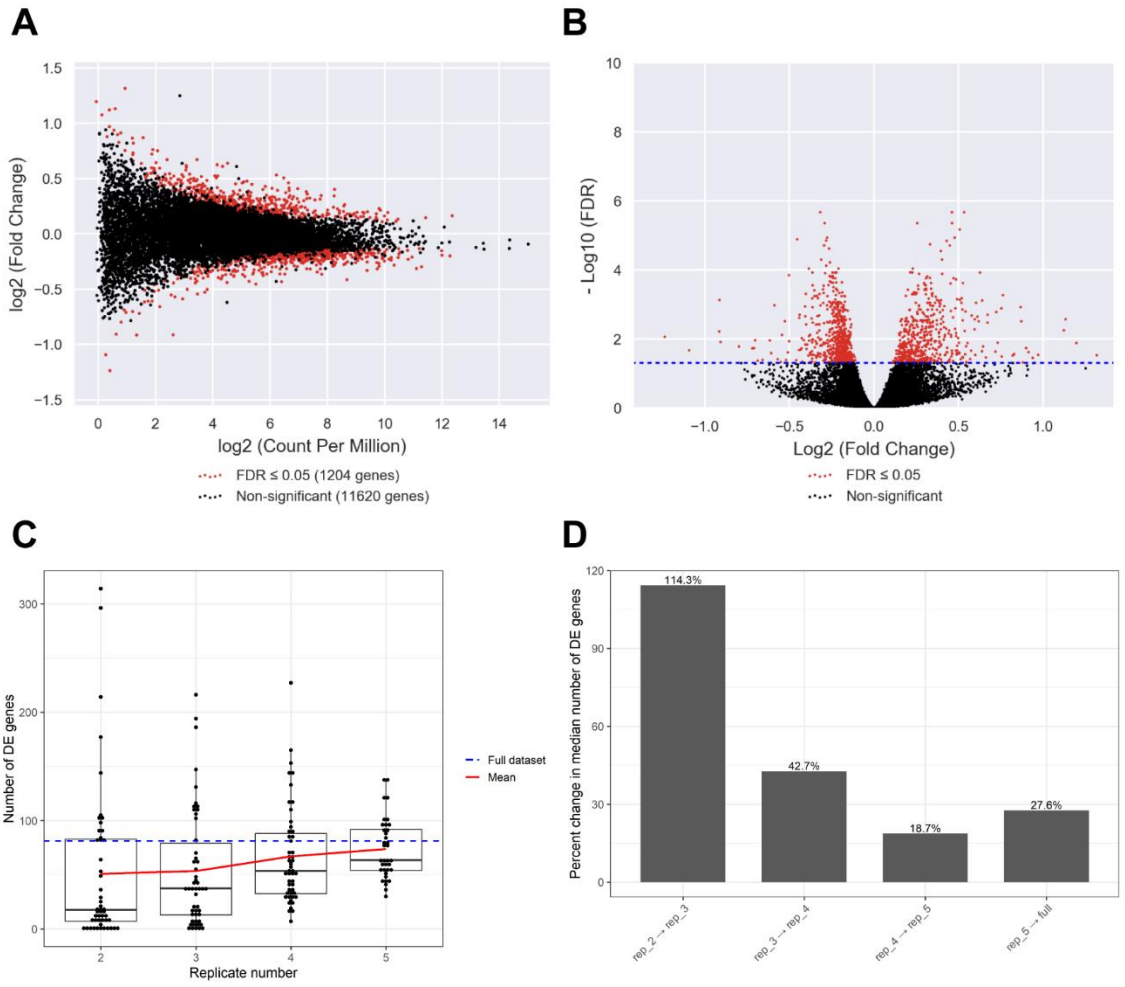

Figure 3.34 *edgeR* comparison of ALG-1001 treatment in PBMC
EdgeR comparison of control and ALG-1001 conditions. **A.** Measured CPM and fold-change for each expressed gene. DE genes are colored red while all others are black. **B.** *edgeR* measured fold-change and FDR displayed for each gene. DE genes are colored red while all others are black. Blue dashed line indicates FDR=0.05. **C-D.** *ERSSA*-generated number of DE gene plot (**C**) and marginal change plot (**D**) with 50 combinations per replicate level and log2-fold-change cut off = 0.5.

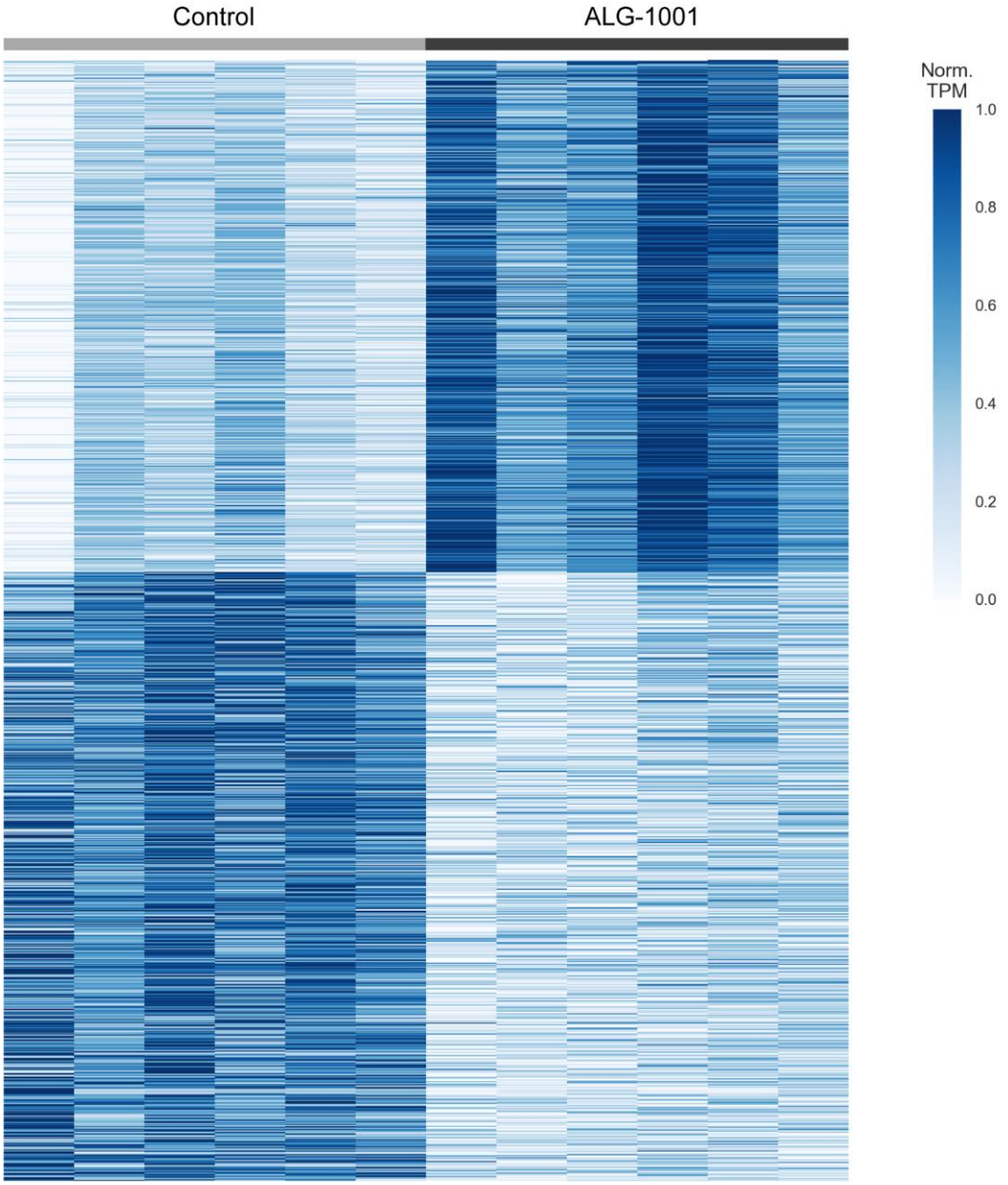

Figure 3.35 Heatmap of 1204 DE genes with ALG-1001 treatment in PBMC
Normalized (Norm.) TPM values are plotted for DE genes between control and ALG-1001 treatment conditions. Genes are plotted on rows and are ranked by decreasing log2-fold-change.

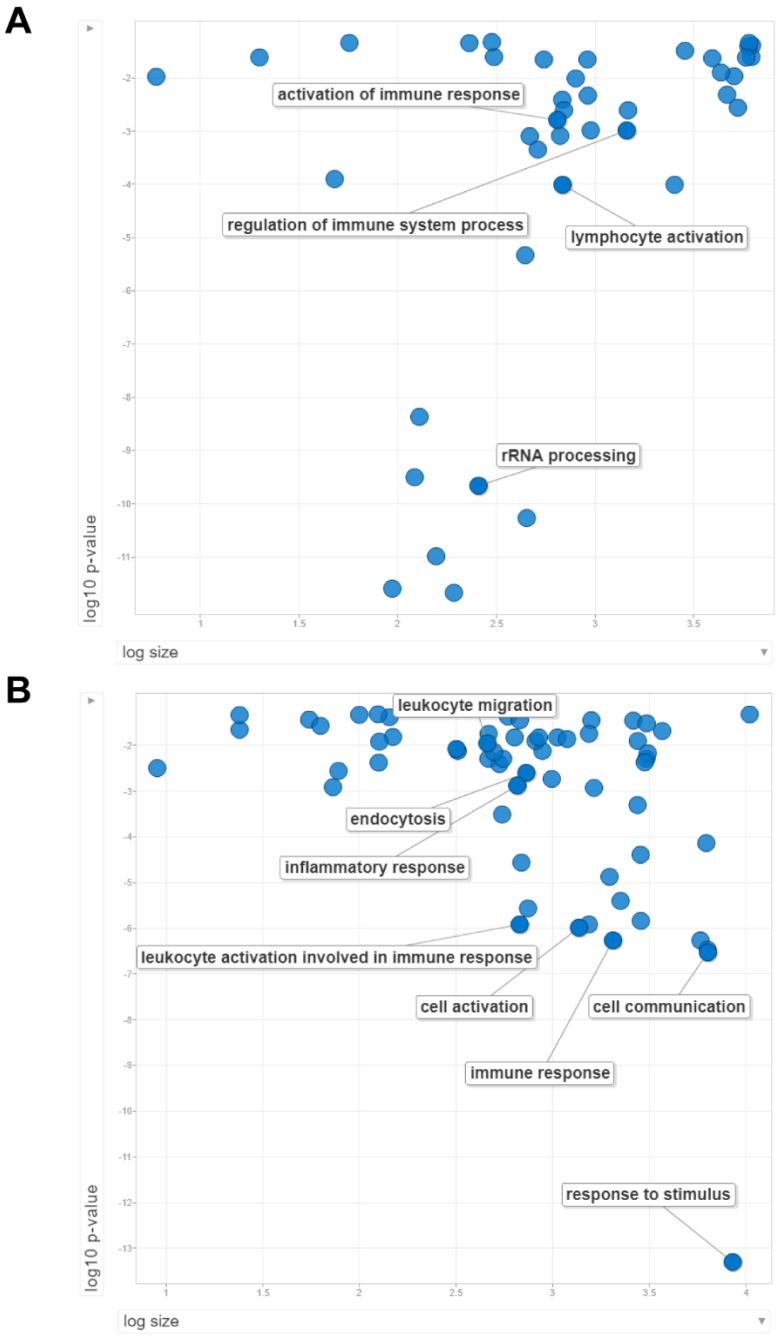

Figure 3.36 GO enrichment of DE genes with ALG-1001 treatment in PBMC
Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO term. Selected biologically relevant GO terms were labeled.

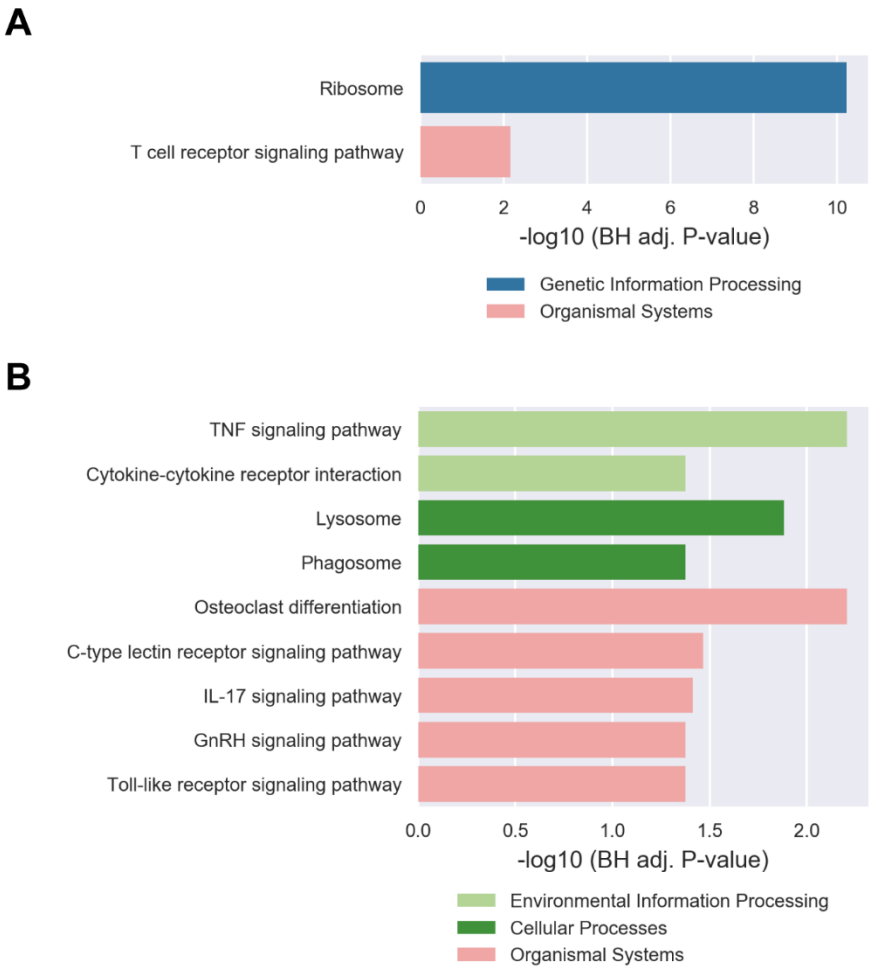

Figure 3.37 KEGG pathway enrichment of DE genes with ALG-1001 treatment in PBMC

KEGG pathways enriched with up- (**A**) or down-regulated (**B**) DE genes using *goseq*. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis.

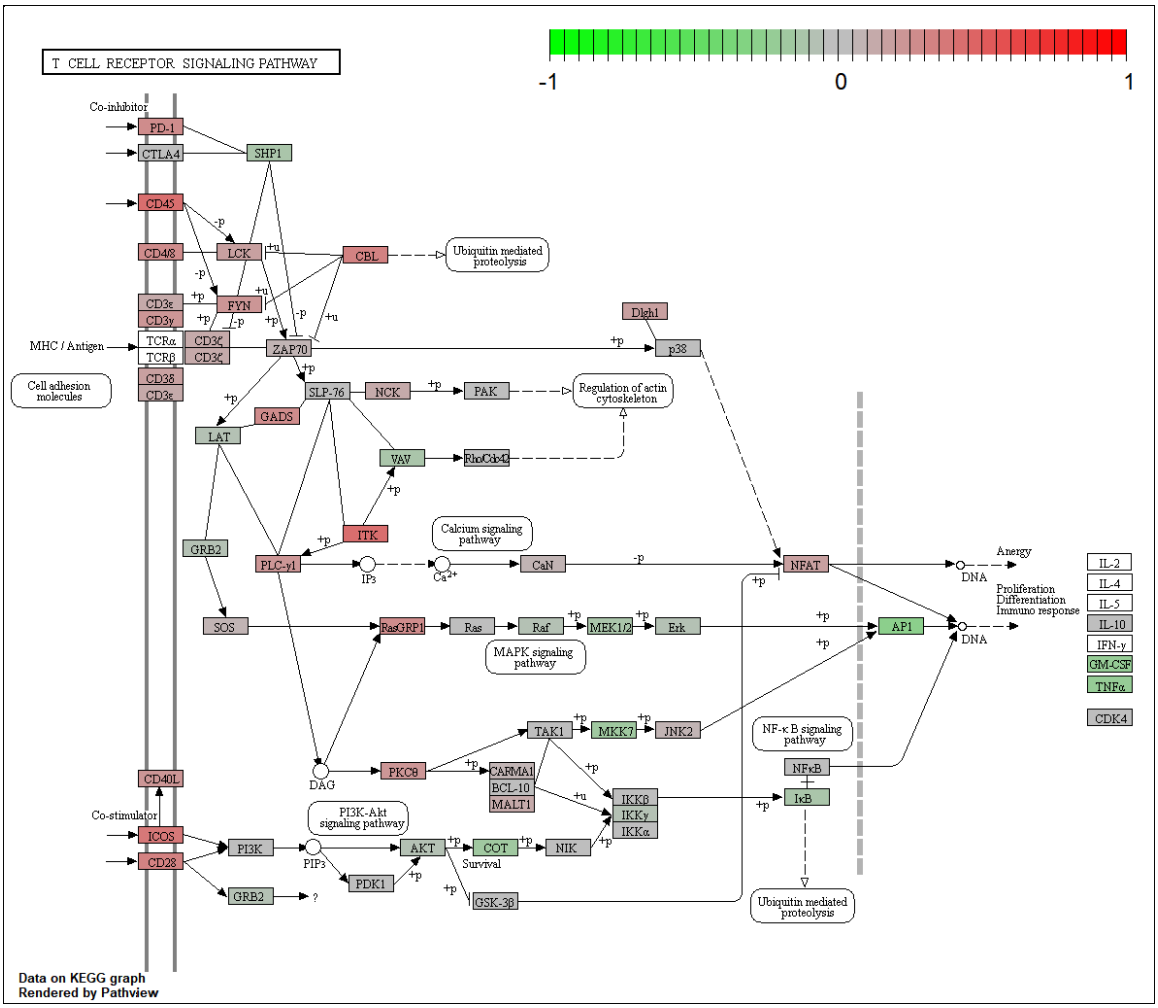

Figure 3.38 *Pathview* visualization of the T cell receptor signaling pathway
ALG-1001 treatment induced pathway expression changes in PBMC are plotted. White colored nodes indicate genes that are not found to be expressed in the dataset.

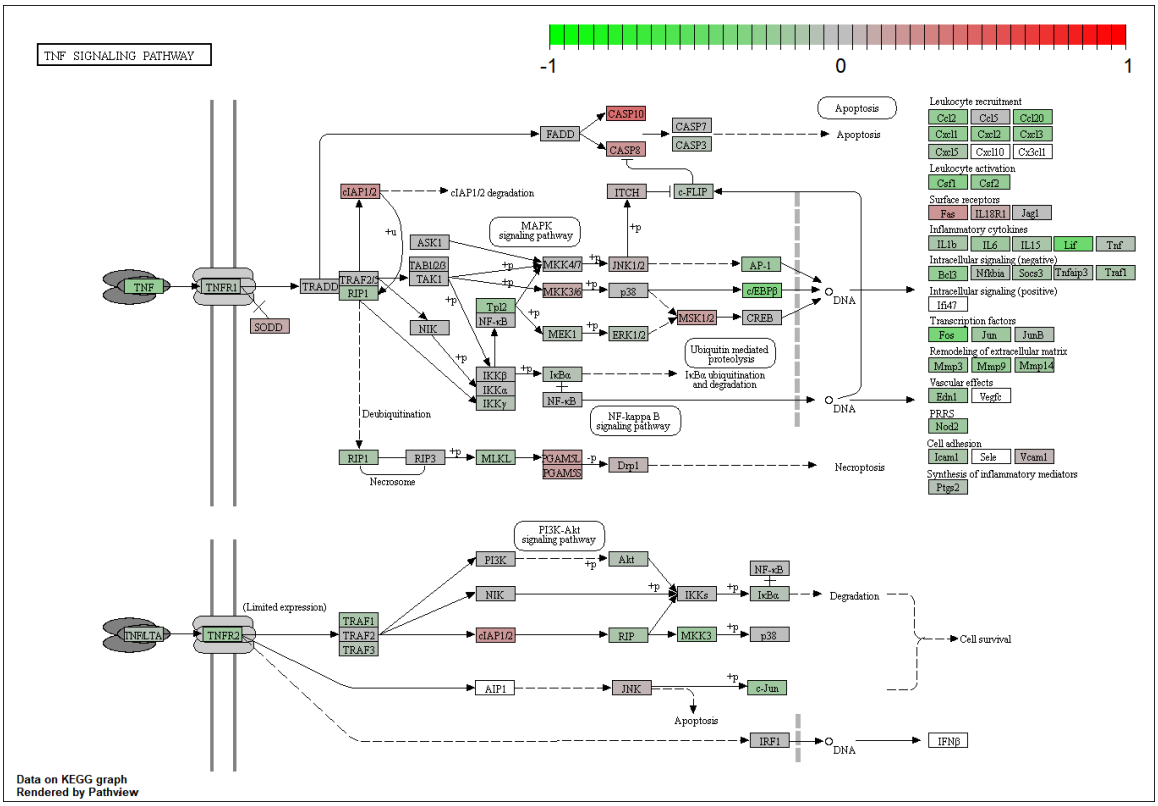

Figure 3.39 *Pathview* visualization of the TNF signaling pathway
ALG-1001 treatment induced pathway expression changes in PBMC are plotted. White colored nodes indicate genes that are not found to be expressed in the dataset.

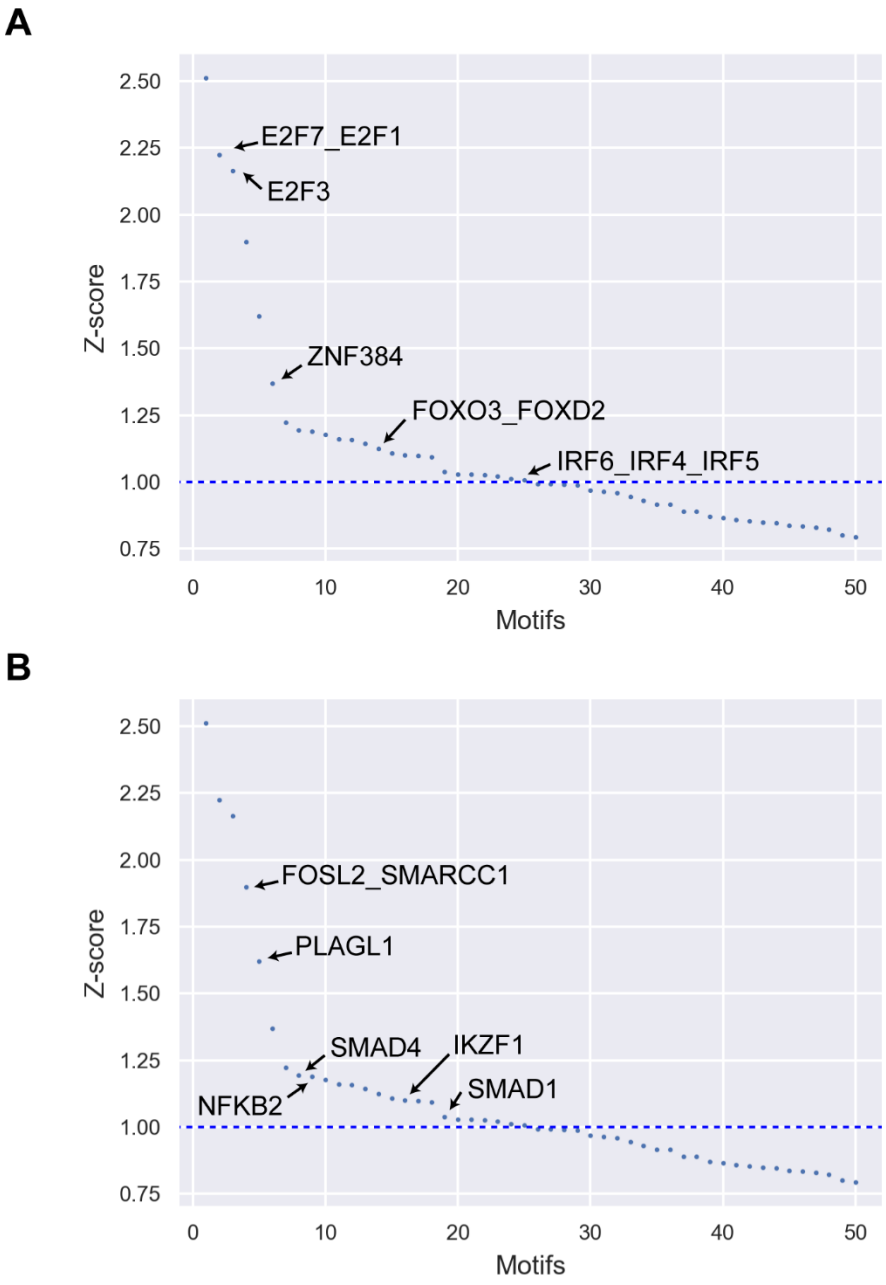

Figure 3.40 *ISMARA* TF enrichment with PBMC samples
PBMC samples were submitted to *ISMARA* for enrichment of TFs that are likely involved in transcriptome regulatory response to ALG-1001 treatment. Top 50 enriched factors with the highest Z-score were plotted (x-axis shows rank by Z-score). Only results with Z-score > 1.0 (blue dashed line) are considered for further analysis. Selected biologically relevant factors with higher (**A**) or lower (**B**) target gene activity in drug treatment condition are labeled.

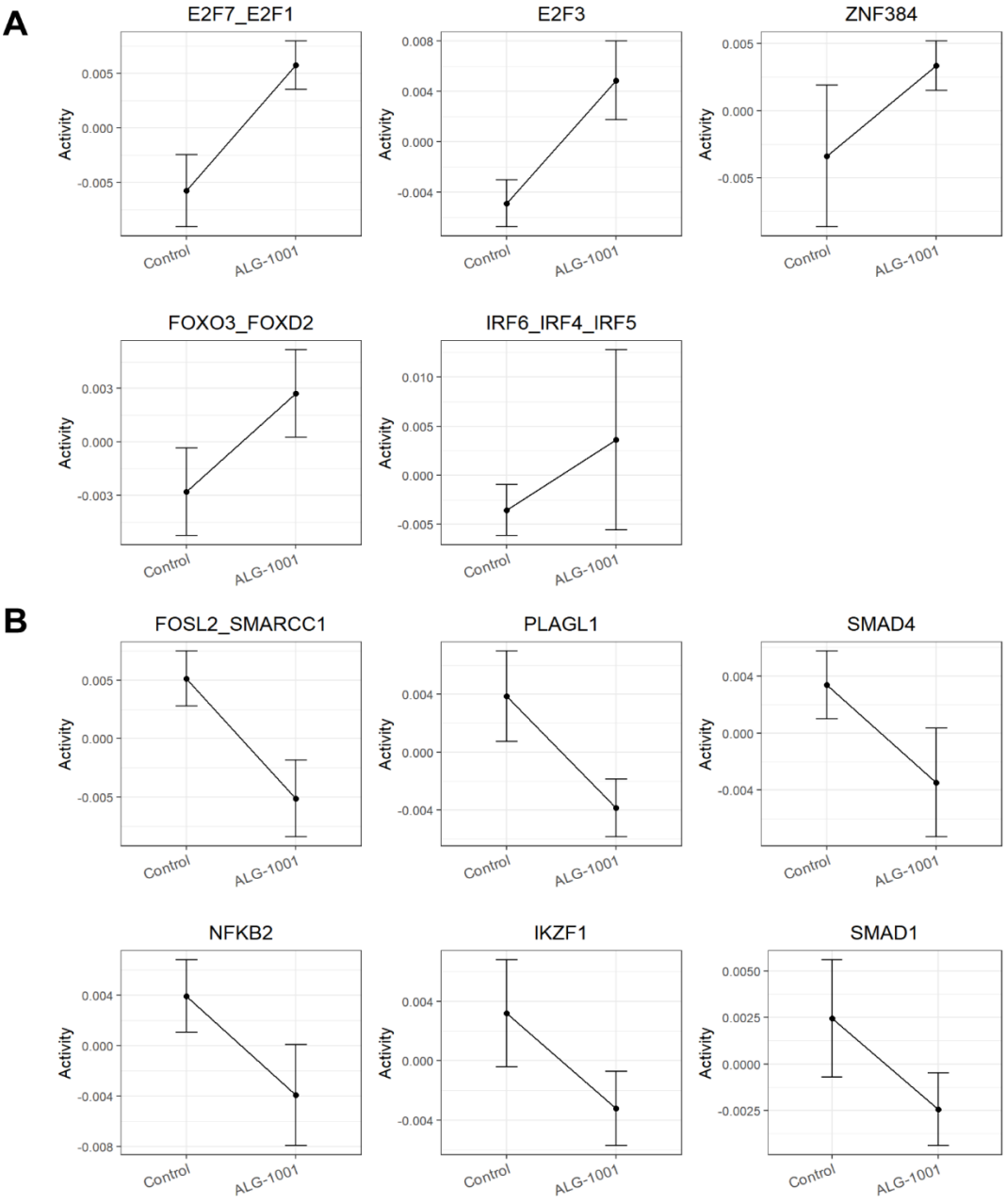

Figure 3.41 Activity profile of selected TF motifs predicted to be regulated by ALG-1001 in PBMC samples

PBMC samples were submitted to *ISMARA* for enrichment of TFs that are likely involved in the transcriptome regulatory response to ALG-1001 treatment. Activity profile (mean + standard deviation) of selected TFs with higher (**A**) or lower (**B**) target gene activity in ALG-1001 condition are plotted.

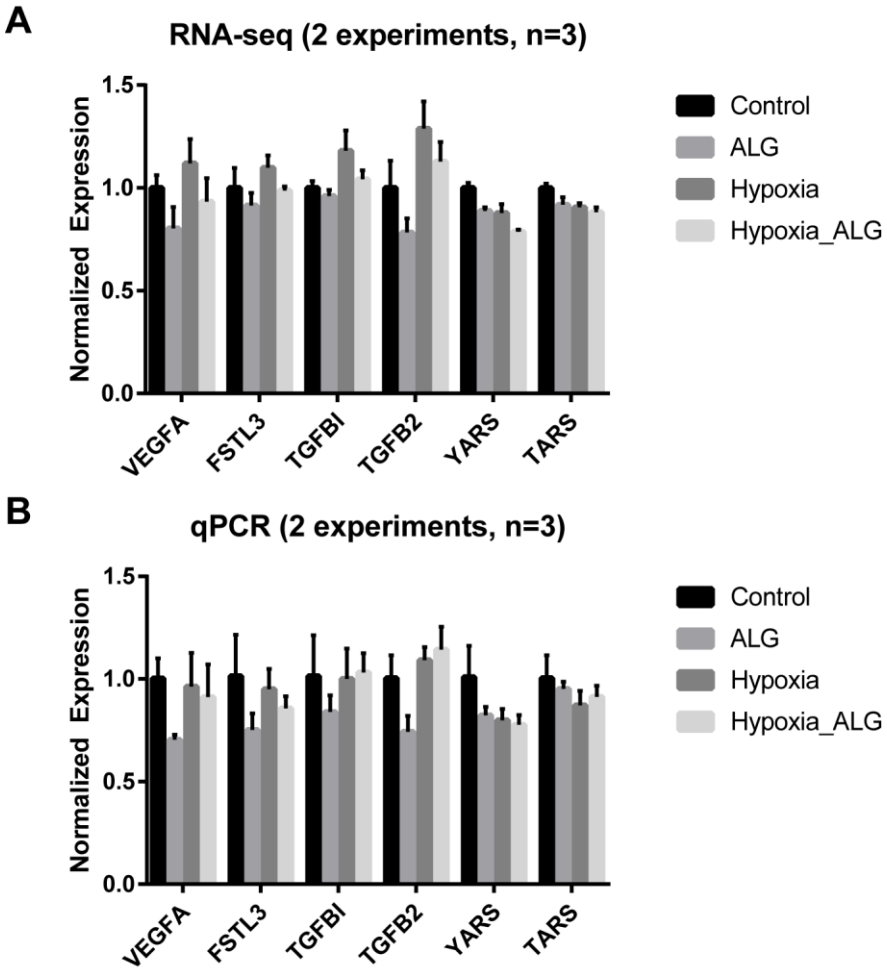

Figure 3.42 qPCR validation of selected HUVEC RNA-seq results
Six genes shown to be regulated by ALG-1001 were selected for qPCR validation. Expression values were normalized in both plots so that average Control expression is 1. The mean and standard deviation are shown. Conditions are labeled as follows: ALG-1001 as ALG, ALG-1001 + hypoxia as Hypoxia_ALG. **A.** RNA-seq expression are based on TPM values. **B.** qPCR expression are based on $\Delta\Delta C_t$-derived fold changes.

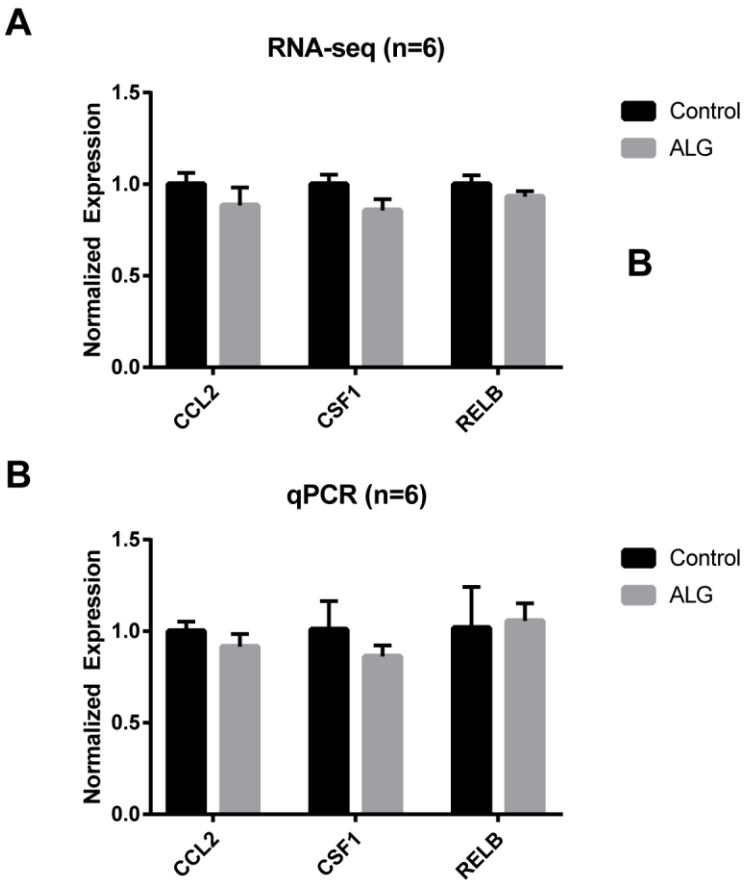

Figure 3.43 qPCR validation of selected PBMC RNA-seq results
Three genes shown to be regulated by ALG-1001 were selected for qPCR validation. Expression values were normalized in both plots so that average Control expression is 1. The mean and standard deviation are shown. ALG-1001 condition is labeled as ALG. **A.** RNA-seq expression are based on TPM values. **B.** qPCR expression are based on $\Delta\Delta C_t$-derived fold changes.

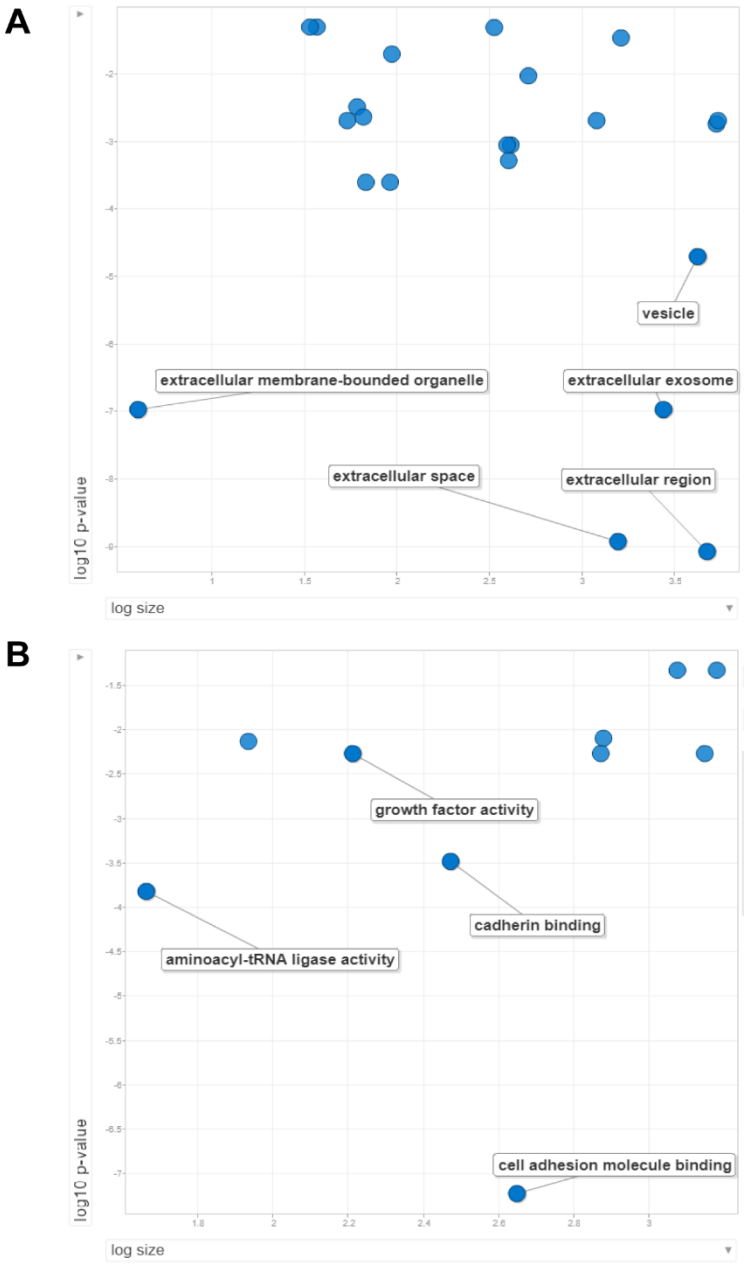

Figure 3.44 GO cellular component and molecular function enrichment of ALG down-regulated genes in HUVEC under hypoxia

ALG-1001 down-regulated DE genes in HUVEC with hypoxia co-streatment were submitted for cellular component (**A**) and molecular function (**B**) GO enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of mouse proteins associated with each GO term. Selected biologically relevant GO terms are labeled.

Gene	Direction	5' to 3' Sequence
VEGFA	Forward	AGGGCAGAATCATCACGAAGT
VEGFA	Reverse	AGGGTCTCGATTGGATGGCA
FSTL3	Forward	ACATTGACACCGCCTGGTC
FSTL3	Reverse	GTCGCACGAATCTTTGCAG
TGFBI	Forward	CACTCTCAAACCTTTACGAGACC
TGFBI	Reverse	CGTTGCTAGGGGCGAAGATG
TGFB2	Forward	CAGCACACTCGATATGGACCA
TGFB2	Reverse	CCTCGGGCTCAGGATAGTCT
YARS	Forward	ACCTCCACGCATACCTGGATA
YARS	Reverse	TTGCTGAGCTGGTAATCAGTG
TARS	Forward	GGAGAAGCCGATTGGTGCT
TARS	Reverse	TCAACTCAGCTCGACCTCCAT
ACTB	Forward	CATGTACGTTGCTATCCAGGC
ACTB	Reverse	CTCCTTAATGTCACGCACGAT
CCL2	Forward	CAGCCAGATGCAATCAATGCC
CCL2	Reverse	TGGAATCCTGAACCCACTTCT
CSF1	Forward	TCCAGTTGCTGGAGAAGGTC
CSF1	Reverse	ACATCTTGGCTGGAGCATTC
RELB	Forward	AGCTACGGCGTGGACAAG
RELB	Reverse	TGGGGGTCAGAGCTGTTC

Table 3.1 qPCR primers

		Hypoxia vs. Control		
		Up	**Down**	**Total**
ALG-1001 + Hypoxia vs. Hypoxia	**Up**	125	139	472
	Down	164	71	450
	Total	2674	2702	

Table 3.2 Two-way contingency table of hypoxia and ALG-1001 with hypoxia treatment regulated DE genes

A larger portion (61%) of commonly regulated genes are in the reverse direction across the two comparisons.

		ALG-1001 vs. Control		
		Up	**Down**	**Total**
ALG-1001 + Hypoxia vs. Hypoxia	**Up**	203	0	472
	Down	0	151	450
	Total	816	654	

Table 3.3 Two-way contingency table of ALG-1001 regulated DE genes

All of the commonly regulated genes are in the same direction across the two ALG-1001 comparisons.

Trans. Factor	Z-score	Activity.diff_ Hypoxia-Control	Activity.diff_ ALG48h-Control	Activity.diff_ Hypoxia_ALG-Hypoxia
SIX5_SMARCC2_HCFC1	3.984283	-0.00416	0.00095	0.000501
TAF1	3.799678	-0.00642	-0.00235	-0.00037
E2F7_E2F1	3.695472	-0.0064	-0.00136	9.50E-05
NRF1	3.647543	-0.005	0.00031	5.00E-04
ELK4_ETV5_ELK1_ELK3_ELF4	3.202669	-0.00497	-0.002	0.000284
TFDP1	3.181176	-0.0051	-0.00095	0.000183
SMAD4	3.043193	0.005763	0.002883	-0.00094
HIC2	3.010955	0.006124	0.000152	-0.00096
WRNIP1	2.937209	0.002543	0.001362	-0.00024
NFIX_NFIB	2.920625	0.005313	0.001506	-0.00077
PATZ1_KLF4	2.862023	0.004281	0.001677	-0.00124
HOXB6_PRRX2	2.860698	-0.00449	-0.00197	-0.0007
ZBTB33_CHD2	2.68979	-0.00538	-0.0024	0.00074
EBF1	2.63914	0.004784	0.000684	-0.00124
SIN3A_CHD1	2.521402	-0.00536	-0.00229	0.000184
TAL1	2.465716	0.002684	0.00262	0.000746
GATA3	2.463018	0.002618	0.001966	-5.00E-06
TBP	2.416656	0.004614	0.001385	0.000431
PLAGL1	2.402684	0.005463	0.001948	-0.00086
POU3F2	2.401511	-0.00412	-0.00151	0.001395
ZBTB14	2.333934	-0.00358	-0.00157	0.000131
STAT5A	2.302745	-0.00381	-0.00241	8.70E-05
RUNX1_RUNX2	2.299625	0.003737	0.000961	-0.00043
POU2F2_POU3F1	2.272958	0.004328	0.001928	-6.70E-05
MYB	2.272434	-0.00373	-0.00117	-0.00018
IKZF1	2.261967	0.003544	0.001164	4.90E-05
AHR_ARNT2	2.255568	-0.0031	-0.00024	2.90E-05
POU5F1_POU2F3	2.210019	0.00378	0.001746	-0.00056
ZNF384	2.190534	-0.00343	-0.00134	0.001379
TFAP2C	2.107671	0.003238	-0.00094	-0.00078
ESRRA_ESR2	2.077387	0.004888	0.000632	-0.00102
MEIS2	2.047777	-0.00362	-0.00086	0.000553

ZFX	2.047151	-0.00289	-0.00121	-0.00048
E2F3	2.030033	-0.00359	-3.10E-05	0.000927
ATF2_ATF1_ATF3	2.030014	-0.0029	-0.00145	0.000442
EPAS1_BCL3	2.006106	0.003194	0.001539	-0.00039
CREB1	1.963439	-0.00377	0.000645	2.70E-05
SMAD1	1.92378	0.00205	0.00064	-0.00076
CXXC1	1.903042	-0.00357	-0.00186	0.000306
SP1	1.898054	-0.00206	0.000159	0.001339
ERG	1.880574	0.002637	0.00184	0.000737
WT1_MTF1_ZBTB7B	1.800999	0.001699	0.000619	-0.00021
ZNF143	1.67804	-0.0028	-0.00125	-0.00041
MECP2	1.673895	-0.00186	0.000171	-0.00072
DLX4_HOXD8	1.667185	-0.00281	-0.00048	0.000587
FOXD1_FOXO1_FOXO6_FOXG1_FOXP1	1.665125	0.002779	0.00085	-0.00037
GMEB2	1.659293	-0.00287	-0.00115	7.20E-05
TLX1_NFIC	1.645351	0.003109	0.000426	-7.30E-05
BARHL1	1.639518	-0.00295	-0.00063	0.001171
ZNF711_TFAP2A_TFAP2D	1.625267	-0.00151	-9.20E-05	4.30E-05

Table 3.4 Top 50 TF motifs enriched with hypoxia treatment in HUVEC

TFs are ordered by Z-score. Activity.diff_Hypoxia-Control is the average target gene activity difference between hypoxia and control conditions. Activity.diff_ALG48h-Control is the average target gene activity difference between ALG-1001 and control conditions. Activity.diff_Hypoxia_ALG-Hypoxia is the average target gene activity difference between ALG-1001 + hypoxia and hypoxia conditions. Bolded TFs indicate ones that are also found on the list of top 50 TFs enriched with ALG-1001 and hypoxia co-treatment.

Trans. Factor	Z-score	Activity.diff_ Hypoxia-Control	Activity.diff_ ALG48h-Control	Activity.diff_ Hypoxia_ALG-Hypoxia
TAL1	2.330915	0.002684	0.00262	0.000746
GATA3	1.581428	0.002618	0.001966	-5.00E-06
KLF16_SP2	1.53227	-0.00017	0.001613	0.000985
TAF1	1.465342	-0.00642	-0.00235	-0.00037
ID4_TCF4_SNAI2	1.449888	0.001203	0.00179	0.000794
STAT5A	1.4191	-0.00381	-0.00241	8.70E-05
SMAD4	1.410515	0.005763	0.002883	-0.00094
WRNIP1	1.406317	0.002543	0.001362	-0.00024
MYF6	1.30968	0.001564	0.002249	0.00125
ERG	1.273625	0.002637	0.00184	0.000737
HOXB6_PRRX2	1.21669	-0.00449	-0.00197	-0.0007
ATF4	1.154007	-0.00086	-0.00243	-0.00187
CEBPB	1.148956	-0.00129	-0.00237	-0.00165
YBX1_FOS_NFYC_NFYA_NFYB_CEBPZ	1.095402	0.000732	0.001037	0.00047
ZBTB33_CHD2	1.026279	-0.00538	-0.0024	0.00074
HOMEZ	1.00642	-0.00164	-0.00191	0.000347
KLF8	1.00128	0.001451	-0.00123	0.000344
POU5F1_POU2F3	0.994232	0.00378	0.001746	-0.00056
ZBTB14	0.991485	-0.00358	-0.00157	0.000131
POU2F2_POU3F1	0.989412	0.004328	0.001928	-6.70E-05
ATF2_ATF1_ATF3	0.988414	-0.0029	-0.00145	0.000442
RXRG	0.978004	0.00036	0.001877	-0.0003
SIX5_SMARCC2_HCFC1	0.976472	-0.00416	0.00095	0.000501
RHOXF1	0.965833	0.002598	0.001662	0.000477
SIN3A_CHD1	0.962837	-0.00536	-0.00229	0.000184
NFKB2	0.93887	0.002021	0.001801	0.000374
EPAS1_BCL3	0.9382	0.003194	0.001539	-0.00039
MLXIPL	0.934386	-0.00055	-0.00131	-0.00039
NR5A2	0.928886	0.000331	0.001681	0.000279
TCF3_MYOG	0.925927	0.001869	0.001701	0.000297
BARX1	0.919222	0.000486	0.001714	-2.50E-05
CPEB1	0.895738	0.001256	0.001702	2.90E-05

E2F7_E2F1	0.893876	-0.0064	-0.00136	9.50E-05
NR1I2	0.888603	0.000676	0.001762	0.000138
ZNF384	0.875025	-0.00343	-0.00134	0.001379
ELK4_ETV5_ELK1_ELK3_ELF4	0.871213	-0.00497	-0.002	0.000284
MXI1_MYC_MYCN	0.861105	-0.00028	0.001389	-0.0015
CREB3L2	0.852152	-0.00051	-0.00172	0.000144
PATZ1_KLF4	0.850309	0.004281	0.001677	-0.00124
SPIC	0.845122	5.60E-05	0.001521	0.000256
ZFX	0.845066	-0.00289	-0.00121	-0.00048
CDC5L	0.840737	0.000397	0.00161	0.001088
NFIX_NFIB	0.834121	0.005313	0.001506	-0.00077
ELF3_EHF	0.821372	-0.00031	-0.00141	-0.00067
HIC1	0.820189	-0.00042	0.001289	0.000245
ETV4_ETS2	0.787745	0.000559	0.001413	0.000763
ESRRB_ESRRG	0.786311	-0.00063	0.001471	0.00035
ZNF524	0.775437	0.001632	0.00137	-0.00061
CXXC1	0.768383	-0.00357	-0.00186	0.000306
IRF2_STAT2_IRF8_IRF1	0.759538	0.000673	-0.00153	0.000411

Table 3.5 Top 50 TF motifs enriched with ALG-1001 treatment under room air oxygen tension in HUVEC

TFs are ordered by Z-score. Activity.diff_Hypoxia-Control is the average target gene activity difference between hypoxia and control conditions. Activity.diff_ALG48h-Control is the average target gene activity difference between ALG-1001 and control conditions. Activity.diff_Hypoxia_ALG-Hypoxia is the average target gene activity difference between ALG-1001 + hypoxia and hypoxia conditions.

Trans. Factor	Z-score	Activity.diff_ Hypoxia-Control	Activity.diff_ ALG48h-Control	Activity.diff_ Hypoxia_ALG-Hypoxia
SOX3_SOX2	1.362482	-5.00E-06	0.000532	0.001711
MNT_HEY1_HEY2	1.243884	-0.00174	-0.00043	-0.00156
ATF4	1.235011	-0.00086	-0.00243	-0.00187
ZNF384	1.220908	-0.00343	-0.00134	0.001379
POU3F2	1.169375	-0.00412	-0.00151	0.001395
SP1	1.119032	-0.00206	0.000159	0.001339
CEBPB	1.105643	-0.00129	-0.00237	-0.00165
MXI1_MYC_MYCN	1.047564	-0.00028	0.001389	-0.0015
PAX5	1.046509	0.000803	-0.00015	-0.00095
PATZ1_KLF4	1.03941	0.004281	0.001677	-0.00124
KLF16_SP2	1.034759	-0.00017	0.001613	0.000985
EBF1	1.002073	0.004784	0.000684	-0.00124
OLIG3_NEUROD2_NEUROG2	0.954093	-0.00118	0.000622	0.001352
HMX3	0.942457	-0.00147	-0.00028	0.001316
MYF6	0.939048	0.001564	0.002249	0.00125
MYBL2	0.93492	-0.0024	-0.00011	0.00131
IRX3	0.904105	-0.00114	-0.00011	0.001283
BARHL1	0.894953	-0.00295	-0.00063	0.001171
FOSL2_SMARCC1	0.874464	0.001586	1.70E-05	-0.00123
ETV1_ERF_FEV_ELF1	0.84078	-9.00E-04	-9.70E-05	-0.00098
MAZ_ZNF281_GTF2F1	0.822072	0.000742	-0.00072	0.000972
HIVEP1	0.787989	0.002172	0.001181	-0.00114
RAD21_SMC3	0.783436	0.00022	-0.00048	-0.00107
HIC2	0.7788	0.006124	0.000152	-0.00096
TAL1	0.772233	0.002684	0.00262	0.000746
FOXK1_FOXP2_FOXB1_FOXP3	0.766034	0.000221	0.000118	0.001037
CDC5L	0.7587	0.000397	0.00161	0.001088
FOXL1	0.757831	-0.00054	-0.00029	0.001006
MECP2	0.756401	-0.00186	0.000171	-0.00072
SMAD1	0.752081	0.00205	0.00064	-0.00076

ELF2_GABPA_ELF5	0.740125	-0.00167	-0.00098	-0.0007
PLAGL1	0.739074	0.005463	0.001948	-0.00086
ID4_TCF4_SNAI2	0.736997	0.001203	0.00179	0.000794
BPTF	0.724676	-0.00089	-0.00064	0.000837
E2F3	0.703794	-0.00359	-3.10E-05	0.000927
SMAD4	0.699577	0.005763	0.002883	-0.00094
FOXD3_FOXI1_FOXF1	0.689003	2.30E-05	0.000115	0.000951
TFAP2C	0.685647	0.003238	-0.00094	-0.00078
GLI3	0.66094	0.001357	-0.0007	-0.00086
SCRT1_SCRT2	0.651946	-0.00016	0.000298	0.000964
FOXO4	0.646831	-0.00048	-0.00081	0.000951
ERG	0.643106	0.002637	0.00184	0.000737
RCOR1_MTA3	0.636119	-0.0006	8.10E-05	0.000344
LEF1	0.614704	-0.00037	-0.00106	0.000864
MBD2	0.593068	-0.00141	-0.00014	0.000855
FOXC1	0.592749	-0.00024	-1.60E-05	0.000875
ZSCAN4	0.591362	-5.40E-05	-0.00023	0.000887
ZBTB33_CHD2	0.590825	-0.00538	-0.0024	0.00074
NFIX_NFIB	0.589195	0.005313	0.001506	-0.00077
ESRRA_ESR2	0.581249	0.004888	0.000632	-0.00102

Table 3.6 Top 50 TF motifs enriched with ALG-1001 treatment with hypoxia co-exposure in HUVEC

TFs are ordered by Z-score. Activity.diff_Hypoxia-Control is the average target gene activity difference between hypoxia and control conditions. Activity.diff_ALG48h-Control is the average target gene activity difference between ALG-1001 and control conditions. Activity.diff_Hypoxia_ALG-Hypoxia is the average target gene activity difference between ALG-1001 + hypoxia and hypoxia conditions. Bolded TFs indicate ones that are also found in the list of top 50 TFs with hypoxia treatment.

Trans. Factor	Z-score	Activity.diff_ALG-Control
BARHL1	2.510727	0.010332
E2F7_E2F1	2.223354	0.011538
E2F3	2.163938	0.009745
FOSL2_SMARCC1	1.897279	-0.01027
PLAGL1	1.619265	-0.00773
ZNF384	1.368142	0.006736
MIXL1_GSX1_BSX_MEOX2_LHX4	1.222901	0.006275
SMAD4	1.194822	-0.00684
NFKB2	1.188247	-0.00785
ZNF524	1.176369	-0.00768
TFAP2C	1.15979	-0.00529
CXXC1	1.158057	0.005567
PPARG	1.144605	-0.00572
FOXO3_FOXD2	1.125123	0.005529
ESRRA_ESR2	1.107849	-0.00636
IKZF1	1.100988	-0.00641
POU3F2	1.099071	0.007951
HOMEZ	1.092696	0.00778
SMAD1	1.038231	-0.00488
ZBTB33_CHD2	1.029819	0.00906
ZNF740_ZNF219	1.029537	0.006702
BPTF	1.02537	0.005366
SP100	1.020871	0.004835
PITX1	1.012363	-0.00747
IRF6_IRF4_IRF5	1.00791	0.007163
SIX5_SMARCC2_HCFC1	0.993107	0.005058
SREBF1_TFE3	0.992486	-0.00618
TAF1	0.989275	0.010782
EBF1	0.98762	-0.00773
ELK4_ETV5_ELK1_ELK3_ELF4	0.968993	0.006617
FOXD1_FOXO1_FOXO6_FOXG1_FOXP1	0.963061	0.005478
ZNF263	0.959227	-0.00573
MAFB	0.945579	0.005795
GCM2	0.929757	-0.00527
RFX7_RFX4_RFX1	0.91573	-0.00463

RUNX3_BCL11A	0.915547	0.005313
TCF7L1	0.890568	0.004693
NR3C1	0.889856	0.00371
ATF7	0.870925	0.00514
ZNF711_TFAP2A_TFAP2D	0.866604	0.003493
HOXC9	0.859728	0.004473
PATZ1_KLF4	0.854418	-0.00217
KLF15	0.848141	0.004259
HSFY2	0.84565	0.006691
PITX3	0.837512	0.004288
SPDEF	0.835104	0.004976
LEF1	0.830758	0.004643
MAFK	0.82314	0.006295
IRF9	0.802357	0.008305
HOXC12_HOXD12	0.794469	0.004031

Table 3.7 Top 50 TF motifs enriched with ALG-1001 treatment in PBMC
TFs are ordered by Z-score. Activity.diff_ALG-Control is the average target gene activity difference between ALG-1001 and control conditions.

3.7 References

1. Leading Causes of Blindness in the U.S. Available at: http://www.nei.nih.gov/health/fact_sheet.asp.
2. Holekamp, N. M. Overview of diabetic macular edema. *Am. J. Manag. Care* **22**, s284–s291 (2016).
3. Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M. & Bhattacharya, S. S. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.* **11**, 273–284 (2010).
4. Pennington, K. L. & DeAngelis, M. M. Epidemiology of age-related macular degeneration (AMD): associations with cardiovascular disease phenotypes and lipid factors. *Eye Vis.* **3**, (2016).
5. Nguyen, Q. *et al.* Management of neovascular age-related macular degeneration: current state-of-the-art care for optimizing visual outcomes and therapies in development. *Clin. Ophthalmol.* 1001 (2015). doi:10.2147/OPTH.S74959
6. Ehlken, C. *et al.* Switch of anti-VEGF agents is an option for nonresponders in the treatment of AMD. *Eye* **28**, 538–545 (2014).
7. Singer, M. A., Kermany, D. S., Waters, J., Jansen, M. E. & Tyler, L. Diabetic macular edema: it is more than just VEGF. *F1000Research* **5**, (2016).
8. Mackel, M. J., Park, J. Y., Karageozian, H. L. & Karageozian, V. H. Integrin Receptor Antagonists and Their Methods of Use. (2013).
9. A Clinical Trial Designed to Evaluate the Safety and Efficacy Study of Luminate® in Inducing PVD (Posterior Vitreous Detachment) in Non-Proliferative Diabetic Retinopathy - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02435862. (Accessed: 11th September 2018)
10. A Safety And Efficacy Study Of ALG-1001 In Human Subjects With Wet Age-Related Macular Degeneration - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT01749891. (Accessed: 11th September 2018)
11. Senger, D. R. & Davis, G. E. Angiogenesis. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
12. Lim, L. S., Mitchell, P., Seddon, J. M., Holz, F. G. & Wong, T. Y. Age-related macular degeneration. *The Lancet* **379**, 1728–1738 (2012).
13. Cheung, N., Mitchell, P. & Wong, T. Y. Diabetic retinopathy. *The Lancet* **376**, 124–136 (2010).
14. Verhoeckx, K., Cotter, P. & European Cooperation in the Field of Scientific and Technical Research (Organization). *The impact of food bioactives on gut health: in vitro and ex vivo models*. (2015).
15. Koss, M. J. *et al.* Proteomics of Vitreous Humor of Patients with Exudative Age-Related Macular Degeneration. *PLoS ONE* **9**, e96895 (2014).
16. Gariano, R. F. & Gardner, T. W. Retinal angiogenesis in development and disease. *Nature* **438**, 960–966 (2005).
17. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).

18. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
19. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
20. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
21. Jiang, H. & Zhan, T. Unit-Free and Robust Detection of Differential Expression from RNA-Seq Data. *Stat. Biosci.* **9**, 178–199 (2017).
22. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
23. Shao, Z. ERSSA: Empirical RNA-seq Sample Size Analysis. *Bioconductor* Available at: http://bioconductor.org/packages/ERSSA.
24. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
25. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
26. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
27. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
28. Tenenbaum, D. KEGGREST: Client-side REST access to KEGG. *R package version 1.20.0* Available at: http://bioconductor.org/packages/KEGGREST/. (Accessed: 26th September 2018)
29. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).
30. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
31. Balwierz, P. J. *et al.* ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.* (2014). doi:10.1101/gr.169508.113
32. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
33. CaltechDATA. *Caltech Library* Available at: https://www.library.caltech.edu/caltechdata. (Accessed: 9th December 2018)
34. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* **25**, 402–408 (2001).
35. Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* **40**, D1144–D1149 (2012).

36. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
37. Costa, C., Incio, J. & Soares, R. Angiogenesis and chronic inflammation: cause or consequence? *Angiogenesis* **10**, 149–166 (2007).
38. Shibuya, M. Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies. *Genes Cancer* **2**, 1097–1105 (2011).
39. Goda, N. *et al.* Hypoxia-Inducible Factor 1α Is Essential for Cell Cycle Arrest during Hypoxia. *Mol. Cell. Biol.* **23**, 359–369 (2003).
40. Krock, B. L., Skuli, N. & Simon, M. C. Hypoxia-Induced Angiogenesis Good and Evil. *Genes Cancer* **2**, 1117–1133 (2011).
41. Kumar, R., Harris-Hooker, S., Kumar, R. & Sanford, G. Co-culture of Retinal and Endothelial Cells Results in the Modulation of Genes Critical to Retinal Neovascularization. *Vasc. Cell* **3**, 1–15 (2011).
42. Mirando, A. C., Francklyn, C. S. & Lounsbury, K. M. Regulation of angiogenesis by aminoacyl-tRNA synthetases. *Int. J. Mol. Sci.* **15**, 23725–23748 (2014).
43. Perez, V. L. & Caspi, R. R. Immune mechanisms in inflammatory and degenerative eye disease. *Trends Immunol.* **36**, 354–363 (2015).
44. van Lookeren Campagne, M., LeCouter, J., Yaspan, B. L. & Ye, W. Mechanisms of age-related macular degeneration and therapeutic opportunities. *J. Pathol.* **232**, 151–164 (2014).
45. Parmeggiani, F. *et al.* Mechanism of Inflammation in Age-Related Macular Degeneration: An Up-to-Date on Genetic Landmarks. *Mediators Inflamm.* **2013**, e435607 (2013).
46. Matteucci, A. *et al.* Primary Retinal Cultures as a Tool for Modeling Diabetic Retinopathy: An Overview. *BioMed Res. Int.* **2015**, e364924 (2015).
47. Al-Gayyar, M. M. & Elsherbiny, N. M. Contribution of TNF-α to the development of retinal neurodegenerative disorders. *Eur. Cytokine Netw.* **24**, 27–36 (2013).
48. Lawrence, T. The Nuclear Factor NF-κB Pathway in Inflammation. *Cold Spring Harb. Perspect. Biol.* **1**, (2009).
49. Zeng, H., Tso, M. O. M., Lai, S. & Lai, H. Activation of nuclear factor-κB during retinal degeneration in rd Mice. *Mol. Vis.* **14**, 1075–1080 (2008).
50. Jiang, N., Chen, X.-L., Yang, H.-W. & Ma, Y.-R. Effects of nuclear factor κB expression on retinal neovascularization and apoptosis in a diabetic retinopathy rat model. *Int. J. Ophthalmol.* **8**, 448–452 (2015).
51. Romeo, G., Liu, W.-H., Asnaghi, V., Kern, T. S. & Lorenzi, M. Activation of nuclear factor-kappaB induced by diabetes and high glucose regulates a proapoptotic program in retinal pericytes. *Diabetes* **51**, 2241–2248 (2002).
52. Cunha-Vaz, J. G. Diabetic macular edema. (1998).
53. A Safety And Efficacy Study Of Alg-1001 In Human Subjects With Symptomatic Focal Vitreomacular Adhesion - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02153476. (Accessed: 11th September 2018)

54. Cines, D. B. *et al.* Endothelial Cells in Physiology and in the Pathophysiology of Vascular Disorders. *Blood* **91**, 3527–3561 (1998).
55. Buckley, C. D. *et al.* RGD peptides induce apoptosis by direct caspase-3 activation. *Nature* **397**, 534–539 (1999).
56. Konisti, S., Kiriakidis, S. & Paleolog, E. M. Hypoxia—a key regulator of angiogenesis and inflammation in rheumatoid arthritis. *Nat. Rev. Rheumatol.* **8**, 153–162 (2012).
57. Sendoel, A. & Hengartner, M. O. Apoptotic Cell Death Under Hypoxia. *Physiology* **29**, 168–176 (2014).
58. Liu, L. & Simon, M. C. Regulation of transcription and translation by hypoxia. *Cancer Biol. Ther.* **3**, 492–497 (2004).
59. Park, J. E. *et al.* Annexin A3 is a potential angiogenic mediator. *Biochem. Biophys. Res. Commun.* **337**, 1283–1287 (2005).
60. Geng, L., Chaudhuri, A., Talmon, G., Wisecarver, J. L. & Wang, J. TGF-Beta Suppresses VEGFA-Mediated Angiogenesis in Colon Cancer Metastasis. *PLOS ONE* **8**, e59918 (2013).
61. Ween, M. P., Oehler, M. K. & Ricciardelli, C. Transforming Growth Factor-Beta-Induced Protein (TGFBI)/(βig-H3): A Matrix Protein with Dual Functions in Ovarian Cancer. *Int. J. Mol. Sci.* **13**, 10461–10477 (2012).
62. Massagué, J., Seoane, J. & Wotton, D. Smad transcription factors. *Genes Dev.* **19**, 2783–2810 (2005).
63. GeneCards - Human Genes | Gene Database | Gene Search. Available at: https://www.genecards.org/. (Accessed: 16th October 2018)
64. Attwooll, C., Denchi, E. L. & Helin, K. The E2F family: specific functions and overlapping interests. *EMBO J.* **23**, 4709–4716 (2004).
65. Varrault, A. *et al.* Identification of Plagl1/Zac1 binding sites and target genes establishes its role in the regulation of extracellular matrix genes and the imprinted gene network. *Nucleic Acids Res.* **45**, 10466–10480 (2017).
66. Huss, J. M., Garbacz, W. G. & Xie, W. Constitutive activities of estrogen-related receptors: Transcriptional regulation of metabolism by the ERR pathways in health and disease. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1852**, 1912–1927 (2015).
67. Naskar, R., Wissing, M. & Thanos, S. Detection of Early Neuron Degeneration and Accompanying Microglial Responses in the Retina of a Rat Model of Glaucoma. *Invest. Ophthalmol. Vis. Sci.* **43**, 2962–2968 (2002).
68. Raskatov, J. A. *et al.* Gene expression changes in a tumor xenograft by a pyrrole-imidazole polyamide. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16041–16045 (2012).
69. McKeown, S. R. Defining normoxia, physoxia and hypoxia in tumours—implications for treatment response. *Br. J. Radiol.* **87**, (2014).

Chapter 4. Hypoxia-induced biological regulation in ARPE-19 cells

4.1 Introduction

The human eye is a sophisticated organ, meticulously organized to enable the visual perception of our surrounding world. Light enters at front of the eye and is detected by the retina, a relatively thin layer tissue consists of mostly retinal photoreceptors and neurons. During this process, the photon signals are converted to electrical impulses that are then transmitted to brain, where they are interpreted[1]. This process is highly energy demanding, which renders the retina one of the most metabolically active tissue in the body[2]. To sustain the high metabolic need, nutrients are delivered to the photoreceptors through a network of capillaries called choroid. Between the choroid and photoreceptors, there is a monolayer of pigmented and supportive cells called the retinal pigment epithelium (RPE)[3]. Despite being only one cell layer thick, the RPE plays several important functions, including light energy absorption, nutrient and waste transportation, retinal metabolic conversion, growth factor secretion, and more[3,4]. These processes are important for both sustaining the visual system and maintaining tissue homeostasis, while any disruption can lead to retinal degeneration, reduced visual perception and in severe cases, blindness.

Age-related macular degeneration (AMD) and diabetic macular retinopathy (DR) are two commonly occurring retinal diseases, where RPE plays a role in the pathophysiology. In AMD, genetic predisposition combined with environmental stress factors cause RPE damage

during the early phase of the disease, which then contribute to the production of pro-angiogenic factors that drive pathological neovascularization[4–6]. Similarly, in diabetic patients with DR, glucose dysregulation is correlated with RPE cell apoptosis and neovascularization[7–10]. The full pathophysiology of these retinal diseases are complex and remain to be explored, but retinal hypoxia is one of the conditions that commonly occur during the early stages of retinal diseases. While the cause of the hypoxia varies, the down-stream responses include activation of inflammation, oxidative stress, angiogenesis, and metabolic regulation[11–13]. Under normal physiological condition, the hypoxia-activated activities help the tissue return to homeostasis. However, under prolonged stress exposure, as typical of retinal diseases, the hypoxia-activated responses can become chronic and eventually cause irreversible damage to the tissue[13,14]. Thus, it has become increasingly clear that persistent retinal hypoxia and its activated processes contribute to dysregulation of RPE cells and disruption of their normal function in maintaining retinal homeostasis.

Current research in hypoxia-activated RPE response is limited, but largely consistent with our general understanding of hypoxia regulation. In an *in vivo* mouse model study, activated hypoxia was shown to commit the RPE to switch from using oxidative phosphorylation to glycolysis for energy production. This metabolic adaptation induced the RPE to hoard glucose and reduce glucose delivery to the photoreceptors, thus causing retinal degeneration[15]. In other studies with *in vitro* ARPE-19 cells, a cell line derived from human RPE tissue, the putative hypoxia response protein HIF-1α was found to be stabilized after hypoxia exposure and contributed to elevated secretion of the potent pro-angiogenic factor –

vascular endothelial growth factor (VEGF)[16–18]. Recently, the high-throughput RNA-sequencing (RNA-seq) technique was used by Jiang and coworkers to study the transcriptome regulation during acute hypoxia exposure (1-4 hours) in ARPE-19 cells[19]. In that study, the authors found acute hypoxia induced detectable up-regulation of apoptosis, metabolism, ribosome, and translation machinery genes and down-regulation of cell cycle genes. To the best of our knowledge, the Jiang study presents the first application of RNA-seq to study the hypoxia-activated expression change in RPE cells, but the short hypoxia exposure used in the Jiang study cannot capture the transcriptome change associated with chronic hypoxia found in the retinal diseases.

In this study, we explored the transcriptome regulation after varying length of low-oxygen exposure by conditioning ARPE-19 cells with short- (8 hours), intermediate- (2 days), and long-term (2 weeks) hypoxia (3% O_2) followed by RNA-seq to identify the differentially expressed genes compared to untreated (maintained at room oxygen tension) control. We are interested in exploring the overall transcriptome change at each time point, as well as similarities and differences in gene expression after varying length of hypoxia exposure. Additionally, a separate set of samples were prepared with exposure to two weeks of hypoxia followed by two weeks of room air oxygen tension. Here, we asked the question whether the cells can recover from long-term hypoxia stimulation, and if so, what is the extent of the recovery. For each condition, we also performed dead and live staining and ROS measurement to identify the treatments' effect on visible cell death and measurable ROS level. Altogether, our study provides a quantitative view of the transcriptome regulation and

the accompanying physiological changes after different length of hypoxia exposure and recovery in RPE cells.

4.2 Materials and Methods

Cell culture

ARPE-19 cells at passage 19 (P19) were purchased from the American Type Culture Collection and grown to P22-24 for all experiments. Cell cultures were kept in a 37˚C humidified incubator with pH controlled by 5% CO_2 and media exchanged twice per week. Initially, DMEM/F12 (Thermo Fisher) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher) and Antibiotic Antimycotic Solution (Sigma-Aldrich) was used to culture the cells. Once cells reach confluency in the assay plate, the FBS concentration was reduced to 2% for the remainder of the experiments. After the switch to low serum, the cells were cultured for six more weeks with plates moved to and from a hypoxia chamber (Biospherix) with O_2 concentration set to 3% and placed inside the cell culture incubator. All cells were collected or assayed at the end of the six weeks culture period after reaching the appropriate period of hypoxia and room air oxygen tension conditioning (see **Figure 4.1** for more details).

ROS assay

For ROS assay, ARPE-19 cells were cultured on 96-well plates (Falcon). To start the assay, cells were washed once with Phosphate-Buffered Saline (PBS, Thermo Fisher) and then

incubated with 2μM H2DCFDA (Thermo Fisher) in cell culture media for 30 minutes at 37˚C. Next the media-dye mixture was replaced with PBS and ROS-correlated fluorescent signal detected on a microplate reader with excitation at 482nm and emission at 520nm. Experiment was repeated once with n=12 per condition. Differences between hypoxia-treated conditions and control were tested using one-way ANOVA with Dunnett multiple testing correction. Statistical analysis was performed in GraphPad Prism 7.00.

Dead and live staining

For dead and live staining, ARPE-19 cells were cultured on 24-well plates (Falcon). The LIVE/DEAD Viability/Cytotoxicity Kit (Invitrogen) was used to simultaneously capture the status of both dead and live cells. At the start of the assay, cells were washed once with PBS before incubation for 30 minutes at 37˚C in the media-dye mixture. At the end of incubation, cells were washed with PBS and imaged on a Zeiss Axiovert 25CFL microscope. FIJI software package was used to merge fluorescent signals into one image with red indicating dead cells and green indicating live cells. Phase contrast images were captured on the same microscopy system with bright-field illumination. A 10x Lysis Buffer (Thermo Scientific Pierce) was used as the lethal control and was added to the wells to working concentration for 15 minute incubation before staining.

RNA-seq sample preparation

For RNA-seq study, ARPE-19 cells were cultured on laminin-coated Transwell filter inserts (Corning) in six-well tissue culture plates (Falcon) as previously described[20]. At end of cell

culture, cells were washed once with PBS without calcium and magnesium (Thermo Fisher) and collected using Accutase (Stemcell Technologies). Total RNA was extracted using RNeasy Mini Kit (Qiagen) and contaminating DNA removed using TURBO DNA-free kit (Thermo Fisher). RNA quality was measured using Bioanalyzer (Agilent Genomics); all samples were found to have a passing quality score of at least 7.20 RNA integrity number. RNA-seq libraries were then prepared using NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs Inc.) and sequenced on the HiSeq 2500 (Illumina) to generate on average 12.5 million, single-end, 100 base pair reads per sample.

RNA-seq analysis

Quality control

FASTQ files generated by RNA-seq were submitted to *FastQC* for quality control. All samples showed good sequencing quality and were submitted for read alignment. Alignment was performed using *Tophat2* to the *hg38* genome and transcriptome references with *Bowtie 2* sensitivity level set to very sensitive[21]. After read alignment, gene counts were quantified using *HTSeq-count* using the intersection-strict model[22]. *FastQC, tophat2* and *HTSeq-count* reports were visualized using *multiQC* software (**Figure 4.2-4.4**)[23].

Principle Component Analysis (PCA)

PCA was used to visualize the high-dimensional dataset. To eliminate null values, all counts in *HTSeq-count*-derived count table were increased by one. Transcript per Million (TPM) values were then calculated to correct for varying sequencing depth and gene length[24]. To

reduce sensitivity to the uncertainty in TPM of low-expressing genes, PCA excluded genes for which the TPM averaged over all samples in the analysis was < 5. Finally, natural log of TPM values was used as input for PCA. From this analysis, no obvious outlier sample was detected.

EdgeR differential expression comparison

In the analysis of differential expression, it is beneficial to exclude genes that are not expressed (they reduce statistical power for identifying differential expression). Here, we retained expressed genes, defined as those with Count per Million (CPM) > 1 in at least five samples (the number of biological replicates per condition). CPM is calculated to normalize for sequencing depth [24]. Around 13,000 genes remain after filtering. Next, the classical approach of *edgeR* was used to perform all of the differential expression comparisons[25]. After statistical test, genes with False Discovery Rate (FDR) ≤ 0.05 were considered as differentially expressed (DE) and further separated into up-regulated and down-regulated gene lists.

Goseq enrichment analysis of DE gene list

Functional analyses of the DE genes were performed with the *goseq* package, which mitigates gene length bias inherent to RNA-seq[26]. Enrichment was performed with both Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) biological pathways[27,28]. GO terms associated with genes were obtained from *biomart* while KEGG pathways were obtained from the KEGG REST server using the *keggrest* package[29,30]. For

each comparison, enrichment was performed separately for the up- and down-regulated gene lists. GO terms were considered enriched with FDR < 0.05 and visualized using *REVIGO*, a visualization tool that aggregates closely-related GO terms, with similarity level set to small and GO term size determined using the *UniPort Homo Sapiens* database[31]. Selected GO terms were labeled on the *REVIGO* plot base on our perception of their biological relevance. KEGG pathways were considered enriched with Benjamini-Hochberg adjusted p-value < 0.05 (p-values are displayed in bar plots). In addition, "Human Diseases" KEGG pathways were removed to improve clarity.

ERSSA

ERSSA was used to check whether the number of biological replicates used in the RNA-seq experiment is sufficient to identify the majority of DE genes[32]. Analysis was performed with absolute log2-fold-change cutoffs set to 0.5 and 50 subsamples at each replicate level.

DE gene heatmap of log2-fold-change values

Heatmaps of TPM value were used to visualize the expression profile of DE genes. TPM values were first calculated from the unadjusted gene counts; then for each gene, TPM was offset such that the minimum is zero and normalized so the maximum is one. DE genes were clustered using k-mean method with k=6 selected based on clustering performance across experimental conditions as well as the marginal improvement in sum squared error (**Figure 4.26A**). For each cluster, associated genes were used for functional analysis using *goseq* as previously described.

ISMARA

ISMARA was used to identify the transcription factors that may be responsible for the observed profiles of gene expression changes in this study[33]. All samples were submitted to *ISMARA* with analysis focused on the top 50 most active motifs as ranked by the enrichment Z-score. For the HIC2 protein, *STRING* was used to visualize the protein associations between the transcription factor targets[34]. Default settings were used for *STRING* network generation.

Data availability

Raw RNA-seq Fastq files will be available at Gene Expression Omnibus once the study has been published[35]. All enrichment tables will be made available at CaltechDATA[36].

Reverse transcription and quantitative polymerase chain reaction (qPCR)

From the DE genes identified by *edgeR*, six genes were selected for separate validation with qPCR: HIF1A (hypoxia response), LDHA (sustains glycolysis), DLD (oxidative phosphorylation), SOD2 (clears ROS), MMP2 and VEGFA (angiogenesis). SuperScript IV Reverse Transcriptase (ThermoFisher) and oligo(dT)20 primer (ThermoFisher) were used to convert polyadenylated mRNA to cDNA. qPCR was performed using iTaq Universal SYBR Green Supermix (Biorad) on LightCycler 480 (Roche) with 45 amplification cycles. All samples were selected for analysis. To enhance accuracy, four technical qPCR replicates of the same cDNA were generated and the median value selected for fold change calculation.

Differences were calculated using the $\Delta\Delta C_t$ method versus *TUBB*[37]. Primers used in this study were obtained from PrimerBank (**Table 4.1**)[38].

4.3 Results

ROS assay

Hypoxia is a major physiological stimulus that activates many intracellular processes including elevation of ROS production[39]. Using a non-fluorescent dye that becomes highly fluorescent after intracellular oxidation, we measured the ROS level in ARPE-19 cells exposed to different length of hypoxia exposure (**Figure 4.1**). Compared to control, ROS level appears unchanged after 8 hours of acute hypoxia, but is weakly elevated after 2 days and 2 weeks of prolonged hypoxia conditioning (**Figure 4.5**). Intriguingly, cells that were allowed to recovery after 2 weeks of hypoxia did not show increased ROS level compared to control.

Dead and live staining

A dual color dye system was used to stain for dead (loss of plasma membrane integrity) and live (active intracellular esterase activity) ARPE-19 cells after hypoxia treatment. Micrograph results show that in all hypoxia conditions, the cell monolayer was maintained and no elevated density of dead cells was observed (**Figure 4.6**). While it is hard to distinguish individual cells, the overall cell morphology also appeared to be unchanged after treatment.

PCA

For each experimental condition, five samples were generated for RNA-seq to probe the transcriptome regulation modulated by each treatment relative to control. PCA was first used to visualize the RNA-seq samples in a low-dimensional space to identify potential outliers and to visualize the separation between conditions. Based on the analysis of the top three principal components (PC1-3 captured a combined 75% of total variance), no obvious outliers were detected (**Figure. 4.7**). The PC1 alone accounted for 47% of total variance in the dataset and clearly separated hypoxia samples from control and hypoxia recovery samples. On PC2, samples that experienced 2 weeks of hypoxia with and without recovery were moderately separated from the other samples. Intriguingly, there appeared to be a trajectory among the hypoxia samples on PC2: 8 hours-treated samples on one end, 2 days-treated samples in the middle and 2 weeks-treated samples on the other end. In addition, hypoxia recovery samples appeared to cluster close to control samples on PC1 and between control and 2 weeks-treated samples on PC2. Without looking at the DE genes, PCA evidence suggest that: 1) there is a detectable transcriptome change after hypoxia exposure, 2) hypoxia's effect on the transcriptome varies based on treatment length, and 3) gene expression partially recovered after long-term hypoxia treatment, but the recovery was not complete.

Transcriptome regulation during hypoxia exposure

EdgeR was used to identify DE genes between control and each of the hypoxia-treated cohorts without recovery (8 hours, 2 days and 2 weeks). Surprisingly, around 6300-6600 DE genes (representing roughly half of the detectable expressed genes) were discovered in each of the comparisons, representing a major shift in the transcriptome after hypoxia treatment (**Figure 4.8**). This transcriptome shift appeared to occur relatively quickly as it was already detectable after 8 hours of hypoxia exposure. To confirm enough biological replicates were included in the comparisons, we used *ERSSA* to simulate DE discovery at different sample sizes. The simulation showed small marginal improvements past four biological replicates in all three comparisons, suggesting five replicates was sufficient to identify a majority of DE genes with |log2-fold-change|>0.5 (**Figure 4.9-4.11**).

With the DE genes, functional analysis was performed to understand the biological processes that are regulated during hypoxia. At all three time points, we see putative hypoxia-response biological processes were enriched with up-regulated DE genes: "blood vessel development", "immune system process", "response to hypoxia", "cell proliferation", "cell death", and others (**Figure 4.12A, 4.14A, 4.16A**). Some of the angiogenesis-related processes enriched include "actin cytoskeleton organization", "extracellular matrix organization", "locomotion", and "anatomical structure morphogenesis". These findings are consistent with previous studies that found RPE under hypoxia stress will respond by promoting angiogenesis[13,16,40]. Enrichment with the down-regulated DE genes generally identified processes involved in metabolism: "oxidative-reduction process", "small molecule metabolism", "electron transport chain", "cellular respiration", and others (**Figure 4.12B,**

4.14B, 4.16B). Down-regulation of genes related to mitochondria-based metabolism is consistent with hypoxia-induced shut down of aerobic respiration[2,40].

To identify variation in enriched GO biological processes among each of the three hypoxia comparisons, each comparisons' top 100 processes (referred to as "top processes") with the strongest statistical evidence were selected and compared (**Figure 4.19**). This analysis showed there are more than 70 common top processes enriched with up-regulated genes in the three hypoxia comparisons (**Figure 4.19A**). On the other hand, the top processes enriched with the down-regulated genes showed less similarity with only 40 found in all three comparisons and 26 others that are shared between the two latter time points (2 days and 2 weeks-treated hypoxia). More importantly, more than 51 (8 hours), 25 (2 days), and 34 (2 weeks) top processes were found in only one of the hypoxia comparisons. A closer look at the 51 unique top processes at 8 hours revealed many that are involved in transcription and translation, such as "RNA processing", "mRNA processing", "ribosome biogenesis", and "translation" (**Figure 4.19B**). On the other hand, the unique top processes enriched at the latter two times points appeared to be more diverse (**Figure 4.20**).

Beyond GO analysis, DE genes were used to identify over-represented KEGG pathways. Across the pathways enriched, many are related to either inflammation ("cytokine-cytokine receptor interaction", "leukocyte transendothelial migration", "chemokine signaling pathway", etc.) or angiogenesis ("MAPK signaling pathway", "VEGF signaling pathway", "PI3k-Akt signaling pathway", "regulation of actin cytoskeleton", and others). Additionally, eight KEGG pathways were consistently found to be enriched (highlighted in **Figure 4.13A**,

4.15A, 4.17), including “ribosome”, “ECM-receptor interaction”, “Cell adhesion molecules (CAMs)”, “HIF-1 signaling pathway”, “Focal adhesion”, and “complement coagulation cascades”, “Hematopoietic cell lineage”, and “axon guidance”.

Pathway enrichment with the down-regulated genes found a combination of “genetic information processing” and “metabolism” pathways after 8 hours of hypoxia treatment (**Figure 4.13B**). In particular, “oxidative phosphorylation” is enriched along with those related to transcription (“Spliceosome” and “RNA transport”) and translation (“Ribosome biogenesis in eukaryotes” and “Proteasome”). As hypoxia prolong to two days and two weeks, we see significantly more enrichment of a variety of metabolic pathways (**Figure 4.15B, 4.18**), ranging from “pyruvate metabolism” to “fatty acid degradation” and “carbon metabolism”. Among the three time points, four pathways are consistently enriched: “oxidative phosphorylation”, “metabolic pathways”, “thermogenesis”, and “proteasome”.

Transcriptome variation after hypoxia recovery

In agreement with PCA, the differential expression discovery with hypoxia recovery samples is smaller in scale compare to the other hypoxia treatment cohorts (**Figure 4.21**). Unlike the other groups, the number of DE genes discovered is reduced roughly by half to around three thousand genes (**Figure 4.21A**). In addition, their fold changes were much closer to zero, with no DE genes of |log2-fold-change|>2. These observations suggest the transcriptome of hypoxia recovery samples is much closer to control condition, a sharp contrast from other hypoxia groups. Additionally, *ERSSA* showed this comparison would benefit from additional

samples, as many of the genes with small change may need additional replicates to be identified with statistical significance (**Figure 4.21D**).

GO biological process and KEGG pathway enrichment found processes and pathways previously enriched in hypoxia treatment groups. Some of the GO biological processes enriched with up-regulated genes include those involved in angiogenesis ("anatomical structure morphogenesis", "extracellular matrix organization", "actin cytoskeleton organization"), inflammation ("immune system process"), "response to hypoxia", "cell proliferation", and "cell death" (**Figure 4.22A**). On the other hand, enrichment with down-regulated genes revealed those involved in metabolism, including "oxidative-reduction process", "carboxylic acid catabolism", and others (**Figure 4.22B**). Consistent with the GO analysis, KEGG pathway enrichment discovered many pathways previously enriched among all of the three hypoxia treatment groups: 6/8 pathways consistently enriched with up-regulated genes and "metabolic pathways" with down-regulated genes (**Figure 4.23-4.24**).

Next, limited to the genes regulated after 2 weeks of hypoxia exposure, we are interested in observing their expression profile after recovery. Intriguingly, we found a majority of these genes have closer to control expression level after recovery (**Figure 4.25**, fold changes after recovery are closer to zero). However, the Pearson's r correlation is 0.7177, suggesting there is still strong correlation to hypoxia response despite two weeks of recovery at control oxygen tension.

Hypoxia-induced transcriptome response across time series

To further understand gene expression variation among the treatment cohorts, we selected DE genes across all comparisons and used k-mean clustering to form six groups of genes with different expression profile (**Figure 4.26B**). Cluster 1 to 3 appear to contain genes that are up-regulated after hypoxia exposure: cluster 1 show highest expression after 2 days of hypoxia, cluster 2 after 2 weeks of hypoxia, and cluster 3 after 8 hours of hypoxia (**Figure 4.27, 4.28A**). Genes in cluster 4 to 6 are generally down-regulated after hypoxia exposure, with cluster 4 showing lowest expression after 2 weeks, cluster 5 consistently across all three hypoxia time points, and cluster 6 after 8 hours of hypoxia (**Figure 4.28B, 4.29**).

For the genes identified in each cluster, GO biological process and KEGG pathway enrichment were performed to identify associated functions. Out of the three clusters with up-regulated DE genes, only cluster 2 generated enrichment results (**Figure 4.27, 4.28A**), which included many of the same biological processes enriched after all three hypoxia treatments. A closer look at several of these enriched GO biological processes shows genes associated are consistently up-regulated after 8 hours and 2 days hypoxia and eventually reach peak measured expression after 2 weeks (**Figure 4.30**).

Enrichment analysis of the three clusters containing down-regulated genes are consistent with earlier functional analysis. In cluster 6, which showed the sharpest down-regulation after 8 hours of hypoxia, mainly transcription and translation processes are enriched (**Figure 4.29B**). Similarly, enrichment of cluster 4 genes, which showed the lowest expression after 2 weeks of hypoxia, identified a diverse array of metabolic pathways that were previously enriched with down-regulated genes after 2 weeks of hypoxia (compare **Figure 4.28B** with

Figure 4.18). Lastly, cluster 5 contains genes sharply down-regulated at all three hypoxia time points (**Figure 4.29A**), which were used to enrich many pathways related to aerobic respiration (*e.g.* "oxidative phosphorylation" and "citrate cycle").

Transcription factor (TF) enrichment

ISMARA was applied to identify TFs that are likely involved during hypoxia regulation. In the top 50 TF groups enriched as ranked by Z-score, we found many of them are well known regulator of biological processes important during hypoxia response (**Figure 4.31** and **Table 4.2**). Some of the target genes of the enriched biologically-relevant TFs have higher expression during hypoxia (**Figure 4.31A**) while others have lower expression (**Figure 4.31B**). Noteworthy TFs with higher target gene expression include known regulators of hypoxia response (HIF1A, ARNT, ARNT2, EPAS1), cell proliferation/apoptosis (KLF4, SRF, FOSL1, FOSL2, ATF4), inflammation (BCL3, CEBPB, CEBPD), and angiogenesis (ZBTB7B) (**Figure 4.32**)[41]. Interestingly, we see a variety of patterns in the target gene activity with some (*e.g.* HIF1A and ERG) that are highly activated during initial hypoxia followed by a gradual decline and others that do not become involved until after at least 2 days of hypoxia (*e.g.* CEBPB and ATF4). Similarly, TFs with lower target expression are enriched and function in hypoxia-regulated biological processes such as cell proliferation/apoptosis (MYC, MXI1, MYCN, FOS, CEBPZ, HCFC1, TAF1, ATF1/2/3), metabolism (NRF1), oxidative stress response (NFE2L2), and hypoxia response (CREB1) (**Figure 4.33**)[41]. As before, a diverse array of target expression profile were detected, including those that initially have the lowest target gene activity that gradually recovered

(*e.g.* MYC, MXI1, MYCN, CREB1, TAF1) and others have target genes more consistently down-regulated throughout hypoxia exposure (*e.g.* NRF1, FOS, CEBPZ).

We also observed many members of the ETS transcription factor family were enriched, including ERG with up-regulated targets and ETV1, ERF, FEV, ELF1, ELF2, GABPA, and ELF5 with down-regulated targets (**Figure 4.32-33**). ETS family is a large family of proteins known to be active in regulating a wide variety of functions including cell migration, cell proliferation, angiogenesis, inflammation, and apoptosis[42]. Indeed, several ETS family members were also previously found to be involved in transcriptome regulation during hypoxia exposure[43,44].

Since *ISMARA* is an unbiased approach to identify TFs that may be involved during hypoxia, it can also be used to discover TFs with unknown regulatory functions. In this dataset, we found one particular TF – HIC2 that was ranked very high on the enrichment result (number 12, Z=2.752), but have only been studied in a few studies unrelated to hypoxia[45–47]. Activity profile of HIC2 shows its target genes are consistently up-regulated by hypoxia across all three treatment groups (**Figure 4.34A**). Intriguingly, *ISMARA* predict HIC2 regulated several genes important for angiogenesis (VEGFA, MMP2, ADM, TGFA, and collagen proteins) and metabolism (LDHA, PKM and GPI) (**Figure 4.34B**)[41]. Furthermore, HIC2 was also enriched with *ISMARA* analysis of HUVECs after exposure to hypoxia (number 8, Z=3.011, **Chapter 3**).

<u>qPCR</u>

qPCR was used to validate RNA-seq measured expression profile. For this purpose, six genes important in hypoxia-response were selected: HIF1A (hypoxia response), LDHA (sustains glycolysis), DLD (oxidative phosphorylation), SOD2 (clears ROS), MMP2 (angiogenesis), and VEGFA (angiogenesis). RNA-seq detected hypoxia-response varies between the six genes: HIF1A was slightly down-regulated, LDHA, MMP2, and VEGFA were significantly up-regulated, while DLD and SOD2 were down-regulated. For all six genes tested, qPCR measured expression profiles (**Figure 4.35B**) are highly consistent with those identified by RNA-seq (**Figure 4.35A**), demonstrating the accuracy of RNA-seq quantification.

4.4 Discussion

In this study, we employed RNA-seq to probe the cellular transcriptomic response after short (8 hours) to long-term (2 weeks) hypoxia in ARPE-19 cells. Based on this dataset, we found many putative hypoxia response pathways are modulated with the regulation both rapid and dynamic. After 8 hours of hypoxia treatment, we already detected differential expression of slightly less than half of the expressed genes. In particular, it is surprising to see many genes up-regulated more than several folds (~10% of up-regulated gene with more than 1 fold change) after such short amount of hypoxia exposure (**Figure 4.8A**). Next, as hypoxia exposure lengthens to two days and two weeks, we are surprised to see that the number of regulated genes detected stayed roughly the same at around half of the expressed transcriptome (**Figure 4.8C, E**). However, further analysis of these genes suggest the identity of regulated genes can differ among the three time points tested.

First, focusing on genes up-regulated by hypoxia, many biological processes are consistently enriched regardless of the hypoxia treatment time (**Figure 4.12A, 4.13A, 4.14A, 4.15A, 4.16A, 4.17**). This is supported by supplementary analysis of the top 100 enriched GO biological processes found in analysis of each hypoxia group, where 70 out of the 100 top processes are found in all three hypoxia cohorts (**Figure 4.19A**). Time series analysis provides another dimension to this observation. By following four processes that are enriched in each of the three hypoxia groups, we found their associated DE gene expression pattern is in fact dynamic (**Figure 4.30**). Thus, although many biological processes are consistently activated by hypoxia regardless of exposure time, the expression level of the associated genes is dependent on amount of low-oxygen exposure.

Focusing on the specific biological processes up-regulated, we found they are predominantly involved in angiogenesis, immune activation and hypoxia response at all three time points. Angiogenesis and inflammation, in particular, are two well-known physiological responses to low tissue oxygen supply: angiogenesis grows new vessels to improve oxygen supply, while inflammation cleans up damaged cells and cell debris generated by the hypoxic stress[48–50]. Both conditions are also extensively reported to be active in retinal diseases and are associated with disease progression[5,13,51,52]. In fact, studies have previously shown that RPE cells may be the main source of the potent angiogenic factor, VEGF, in the diseased retina[3]. Collectively, these findings suggest hypoxic stress can induce RPE cells to activate disease-relevant physiological processes, which can contribute to disease development.

Next, focusing on down-regulated genes, we found GO biological process and KEGG pathway analyses both suggest transcription and translation-related genes are disproportionally down-regulated after eight hours of hypoxia (**Figure 4.12B, 4.13B, 4.19B**). As some of the most energy intensive processes, mRNA synthesis and protein translation are known to be rapidly suppressed during hypoxia response at a time when energy supply is disrupted[53,54]. Yet, as hypoxia continues, the biological processes that are most highly enriched are replaced with ones that are involved in metabolism, particularly ones involved in aerobic respiration, *i.e.* TCA cycle and oxidative phosphorylation (**Figure 4.14B, 4.15B, 4.16B, 4.18**). This is consistent with our current understanding of hypoxia, where a lack of oxygen supply induces cells to primarily use oxygen-independent glycolysis for energy production. DE gene expression clustering analysis provided additional evidence to support these finding. In the three clusters that contained down-regulated genes after hypoxia exposure, cluster 6 showed the lowest expression after 8 hours of hypoxia treatment and was highly enriched with genes involved with transcription and translation (**Figure 4.29B**). Similarly, cluster 4 and 5 both mainly enriched metabolic related processes with genes down-regulated the most after 2 weeks hypoxia (**Figure 4.28B**) or consistently down-regulated across all hypoxia treatments (**Figure 4.29A**).

In relation to the Jiang *et al.* study that used RNA-seq to measure ARPE-19 transcriptomic response after acute hypoxia exposure (4 hours and less)[19], our results show major similarities as well as differences. First, both datasets found acute hypoxia induced regulation of transcription and translation processes. However, the enrichment results differ from there

as the Jiang study also found activation of apoptosis and suppression of cell cycle genes. Among these two processes, we found very weak enrichment of "cell death" and no functional analysis results that suggest cell cycle down-regulation. We suspect this discrepancy might be due to a relatively mild 3% O_2 hypoxia treatment used in our study compare to the 1% O_2 used in the Jiang study. Consistent with the lack of apoptosis-related biological enrichment, dead and live staining as well as ROS assay did not reveal any systematic loss of cell density or significantly elevation of ROS level after 3% O_2 hypoxia exposure (**Figure 4.5-6**). In a second major difference, the Jiang study did not detect the activation of angiogenesis and inflammatory processes and suppression of metabolic genes. We suspect this might be due to a lack of statistical power in the original study (2 replicates per condition), where only 332 DE genes were identified (versus ~6000 in each of our hypoxia cohorts). As a result, the limited replicates employed likely failed to capture many of the real changes between the conditions.

In our dataset, we also found genes normally associated with "axon guidance" KEGG pathway are up-regulated by hypoxia in ARPE-19 cells. This observation is confounding as a literature search failed to find a connection between RPE cells and axon guidance, which is a key process in the formation of a neuronal network. Intriguingly, a yet unpublished study also found highly statistical significant enrichment of axon guidance pathway, but in the context of RPE cells stimulated to undergo EMT[55]. In that study, 40 axon guidance-related genes were found to be regulated after EMT induction and out of those, 37 are also regulated by hypoxia in at least one of the hypoxia time points in our study. These results suggest genes

normally involved in axon guidance may also be involved in EMT and that regulation of these same genes by hypoxia indicate possible induction of EMT in ARPE-19 cells. Since RPE cells that undergo EMT have been reported in a variety of retinal diseases including AMD[56], our transcriptome findings suggest hypoxia may be one of the contributing factors to the transition. Based on these observations, further investigation in the connection between hypoxia and EMT is warranted, particularly in the context of the possible role EMT may play in retinal disease pathophysiology.

In comparison to our other studies of hypoxia-induced transcriptome response in OIR mice (**Chapter 2**) and human endothelial cells (**Chapter 3**), many of the same biological processes were found to be enriched in ARPE-19 cells. In all three studies, we found inflammation, angiogenesis, and hypoxia-response processes were enriched with up-regulated genes, which suggest activation of these processes is likely to be more universal. On the other hand, enrichment of down-regulated genes produced different results, where the mice study (**Chapter 2**) mainly found down-regulation of neuronal and metabolic processes, while the endothelial study (**Chapter 3**) found down-regulation of both cell cycle and metabolism. Additionally, since both previous studies involved hypoxia exposure longer than 8 hours (5 days in OIR mice and 1 day in endothelial cells), down-regulation of transcription and translation-related genes were not strongly detected.

In this study, we also conditioned ARPE-19 cells to two weeks of hypoxia followed by two weeks of recovery at room air oxygen tension. These hypoxia recovery samples showed expression profile that is fairly close to control samples while retaining some signature of the

hypoxia-induced transcriptome modulation. First, PCA shows hypoxia recovery samples clustered closely with control samples on the major PC that separates the control and hypoxia samples (**Figure 4.7A**). Additionally, PC2 shows recovery samples are situated between control and 2 weeks hypoxia-treated samples, which suggest recovery samples likely have an expression profile between these two conditions. Similarly, DE gene heatmap shows gene expression of hypoxia recovery samples largely resided between control and week 2 hypoxia treated levels (**Figure 4.26**). Compared to the other three time points, comparison of hypoxia recovery samples to control also identified around half as many DE genes, many of which have relatively small fold change and none with |log2-fold-change|>2 (**Figure 4.21A**). Functional analysis of these DE genes revealed many of the same biological processes previously enriched with the other hypoxia treatment conditions, suggesting some of the hypoxia modulated signature is retained despite the long recovery period (**Figure 4.22-4.24**). In short, these results demonstrate ARPE-19 gene expression is adaptable to oxygen level, but some signature of hypoxia remains even after recovery.

Transcription factor enrichment largely supports results of functional analysis based on DE genes (**Figure 4.31**). Focusing on the top 50 enrichment TFs, we recovered many that are known to regulate the biological processes enriched with the hypoxia-regulated genes. In addition, we also enriched a mostly uncharacterized transcription factor – HIC2, with strong statistical significance. Based on *ISMARA* calculation, expression of predicted HIC2 target genes were consistently up-regulated at all three hypoxia time points (**Figure 4.34**) and some of the possible targets include angiogenesis- and metabolic-related genes. At the time we

performed this analysis, less than five HIC2-associated papers have been published and none of them in context of hypoxia activation. Based on these observations, it may be interesting to further explore HIC2 in context of hypoxia response and identify the genes modulated by the transcription factor.

4.5 Conclusions

In the present study, we exposed human RPE cells to three different length (8 hours, 2 days, and 2 weeks) of 3% O_2 hypoxia treatment and probed the hypoxia-induced transcriptome regulation using RNA-seq. Our results suggest the hypoxia-induced transcriptome regulation is both rapid and dynamic with many processes involved, such as activation of angiogenesis and inflammation genes as well as suppression of metabolic genes. Another set of samples prepared with 2 weeks of hypoxia exposure followed by room air oxygen tension recovery showed their transcriptome was in a state between the control and long-term hypoxia treated cells, suggesting at least a portion of the transcriptome response to hypoxia is reversible. We also discovered a poorly studied transcriptome factor that may be involved in hypoxia regulation and found evidence that suggest hypoxia exposure may have activated EMT in RPE cells, both of which are interesting areas for future research.

4.6 Figures and Tables

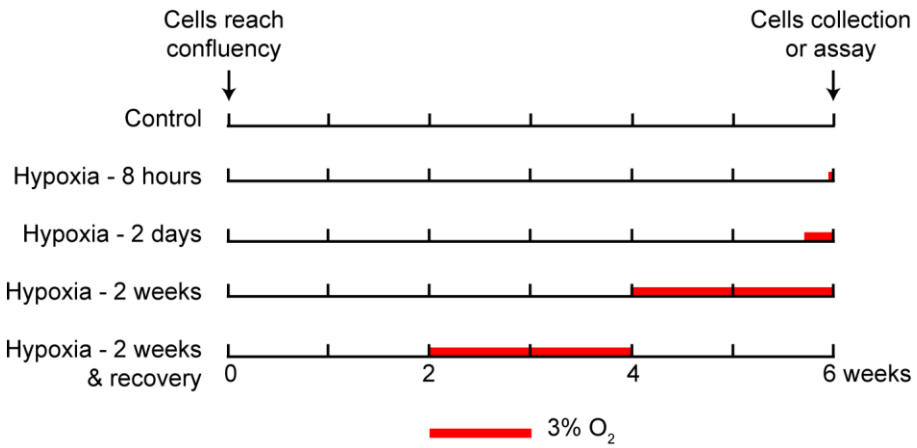

Figure 4.1 APRE-19 hypoxia treatment design
ARPE-19 cells were exposed to 3% O_2 (red bar) or room air oxygen tension (empty) for the specified amount of time before assay at the end of six weeks of confluent cell culture. After reaching confluency (week 0), cells were cultured for two weeks without any hypoxia treatment to induce differentiation into RPE morphology.

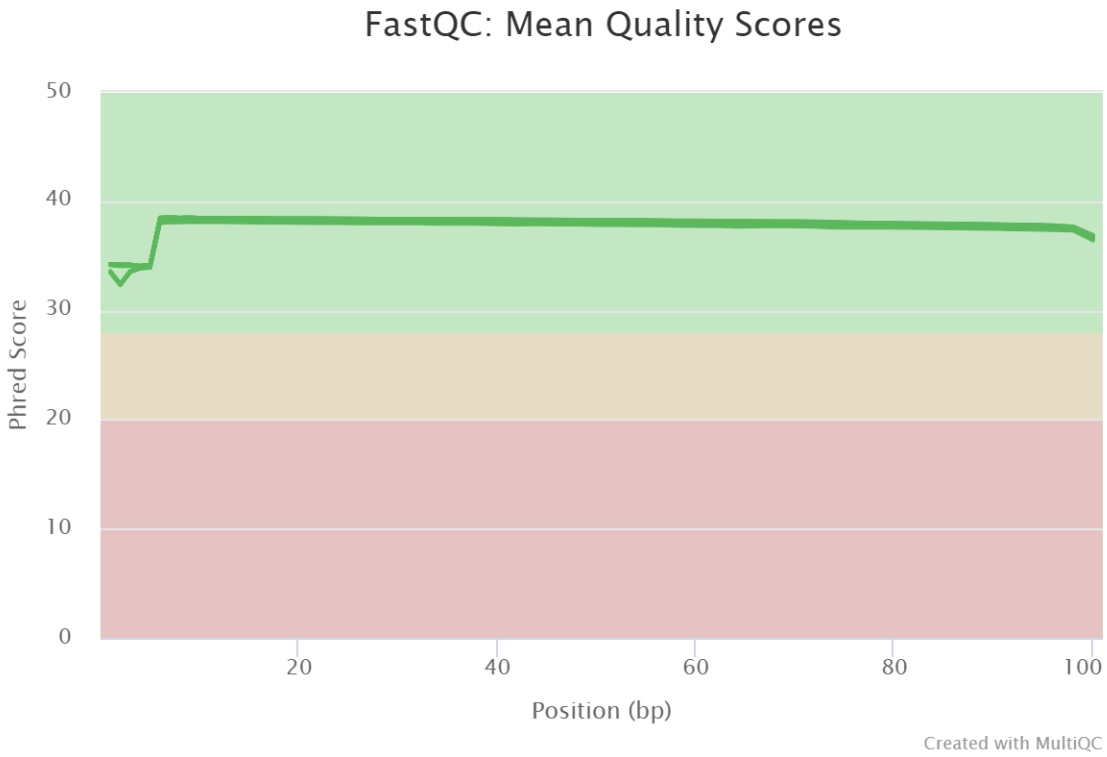

Figure 4.2 *FastQC* mean sequencing quality score
MultiQC plots the average sequencing quality Phred score at each base in the reads generated for each sample. Each line detonates a sample with majority of them overlapping throughout the entire 100 base pair region. Green Phred score indicates good sequence quality while red indicates poor quality.

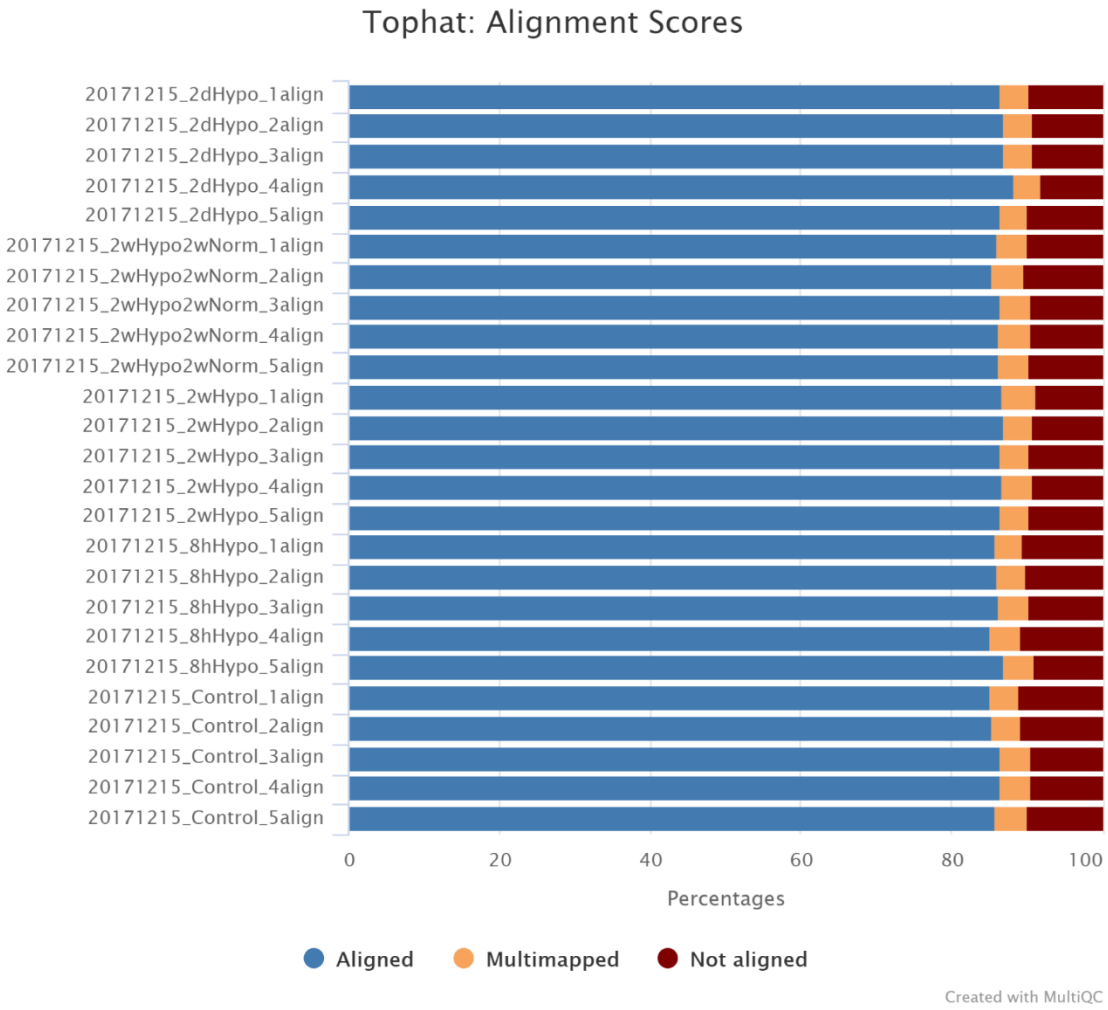

Figure 4.3 *Tophat2* statistics
At least 89% of reads were mapped in all 25 samples by *Tophat2*.

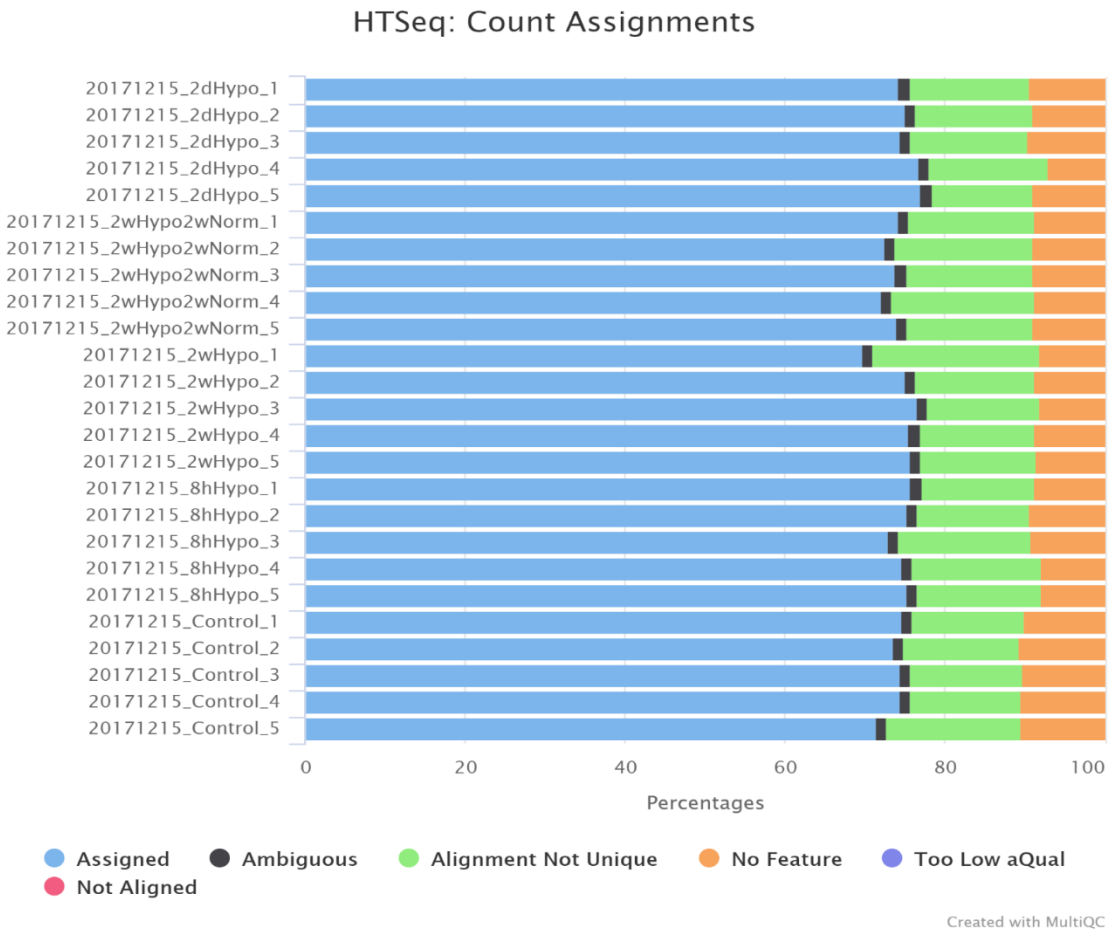

Figure 4.4 *HTSeq-count* statistics
HTSeq-count uniquely assigned (blue bar) between 70-77% of a sample's aligned reads to a gene.

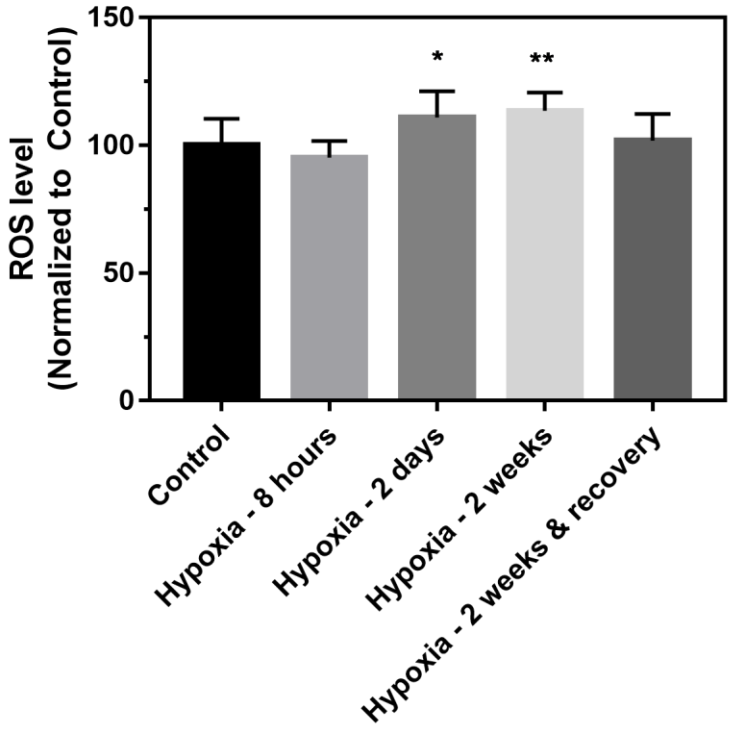

Figure 4.5 ROS level measured after hypoxia treatment
After exposure to varying length of hypoxia treatment, ARPE-19 cells were assayed to measure ROS level using a dye that becomes highly fluorescent after intracellular oxidation. The results suggest there is a weakly elevated ROS level after 2 days and 2 weeks of hypoxia treatment, but no change after 8 hours of hypoxia and after recovery from hypoxia. ROS level were normalized to average measurement in control samples. Mean and standard deviation (n=12) plotted. All statistical tests were performed against control samples with the stars indicating the magnitude of the adjusted p-value (* $P \leq 0.05$, ** $P \leq 0.01$).

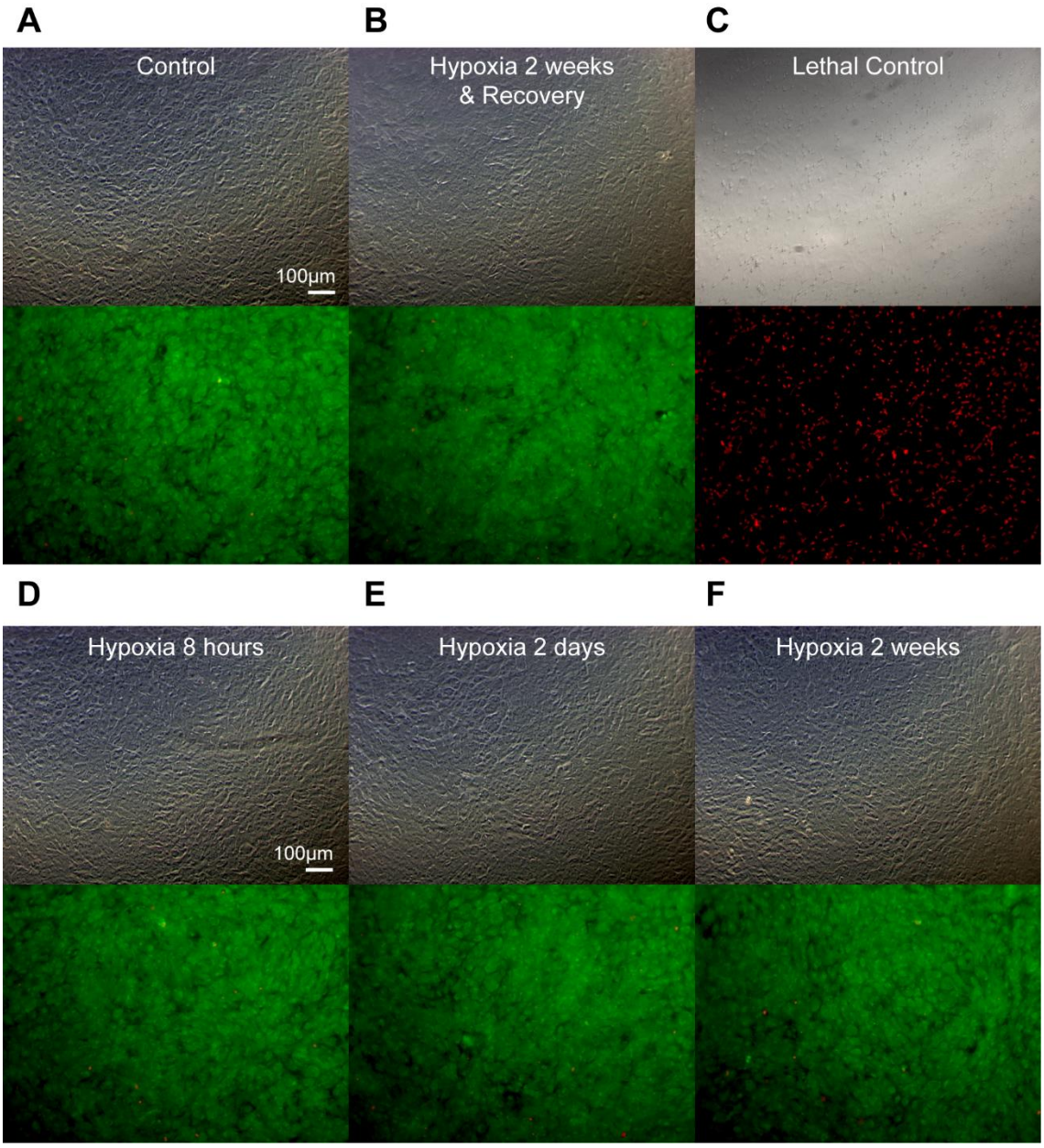

Figure 4.6 Dead and live staining of ARPE-19 after hypoxia treatment
After exposure to varying length of hypoxia treatment (**A.** no hypoxia, **B.** 2 weeks hypoxia and 2 weeks recovery, **D.** 8 hours hypoxia, **E.** 2 days hypoxia, **F.** 2 weeks hypoxia), ARPE19 cells were assayed to stain for dead (red) and live (green) cells as shown in bottom image of each subsection. Corresponding phase contrast micrographs are shown in the top image of each subsection. No elevated cell death is observed after hypoxia treatment. Detergent-treated lethal control is shown in **C.**

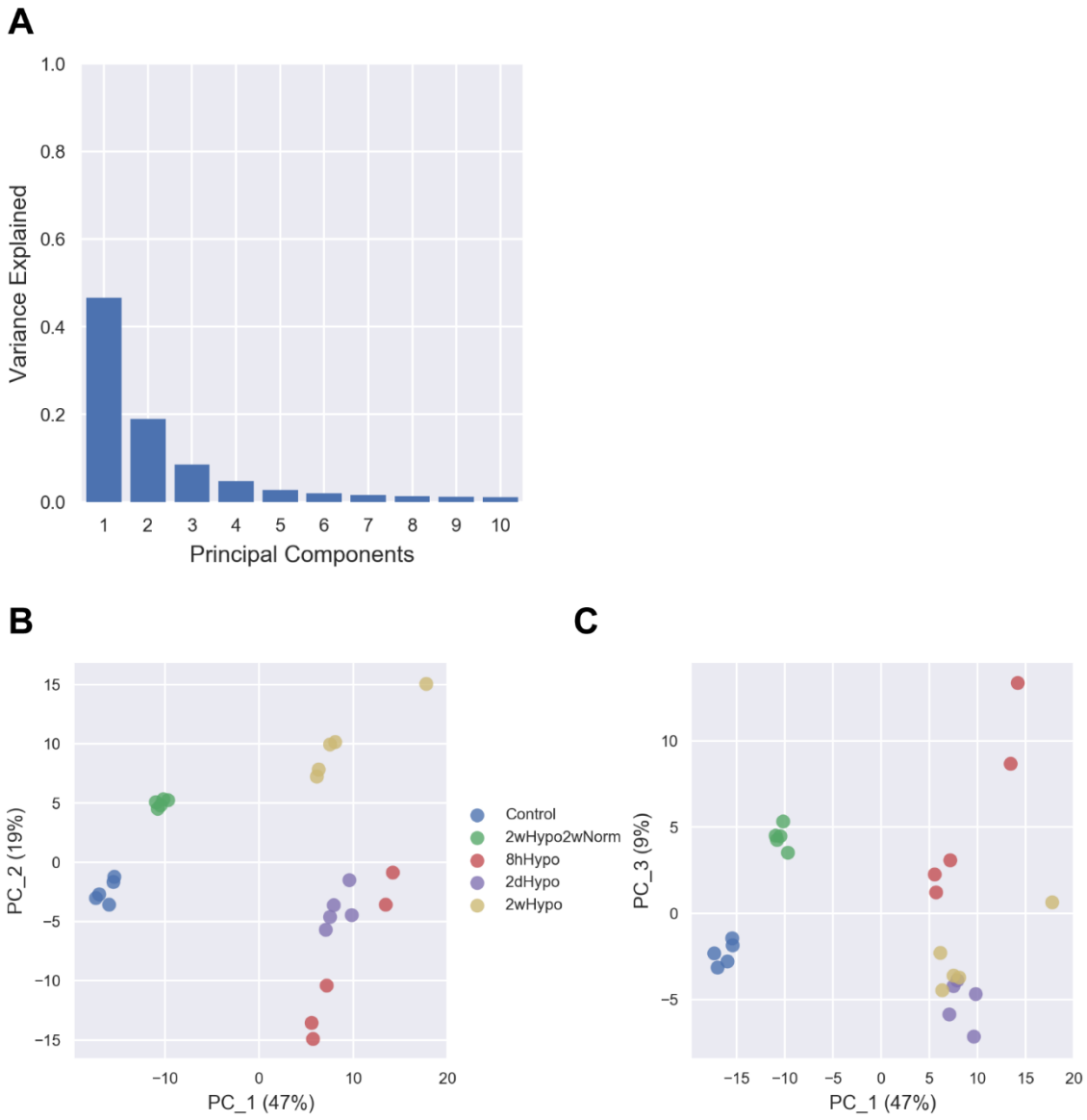

Figure 4.7 PCA of all samples

PCA performed with all 25 ARPE-19 samples. 2wHypo2wNorm is 2 weeks hypoxia followed by recovery, 8hHypo is 8 hours hypoxia, 2dHypo is 2 days hypoxia, and 2wHypo is 2 weeks hypoxia samples. PC # is principal component #. **A.** Percent of overall variance captured by the top 10 principal components. **B-C.** Samples in PC 1 and PC 2, PC 1 and PC 3, respectively. Percent of variance explained is shown in axis label.

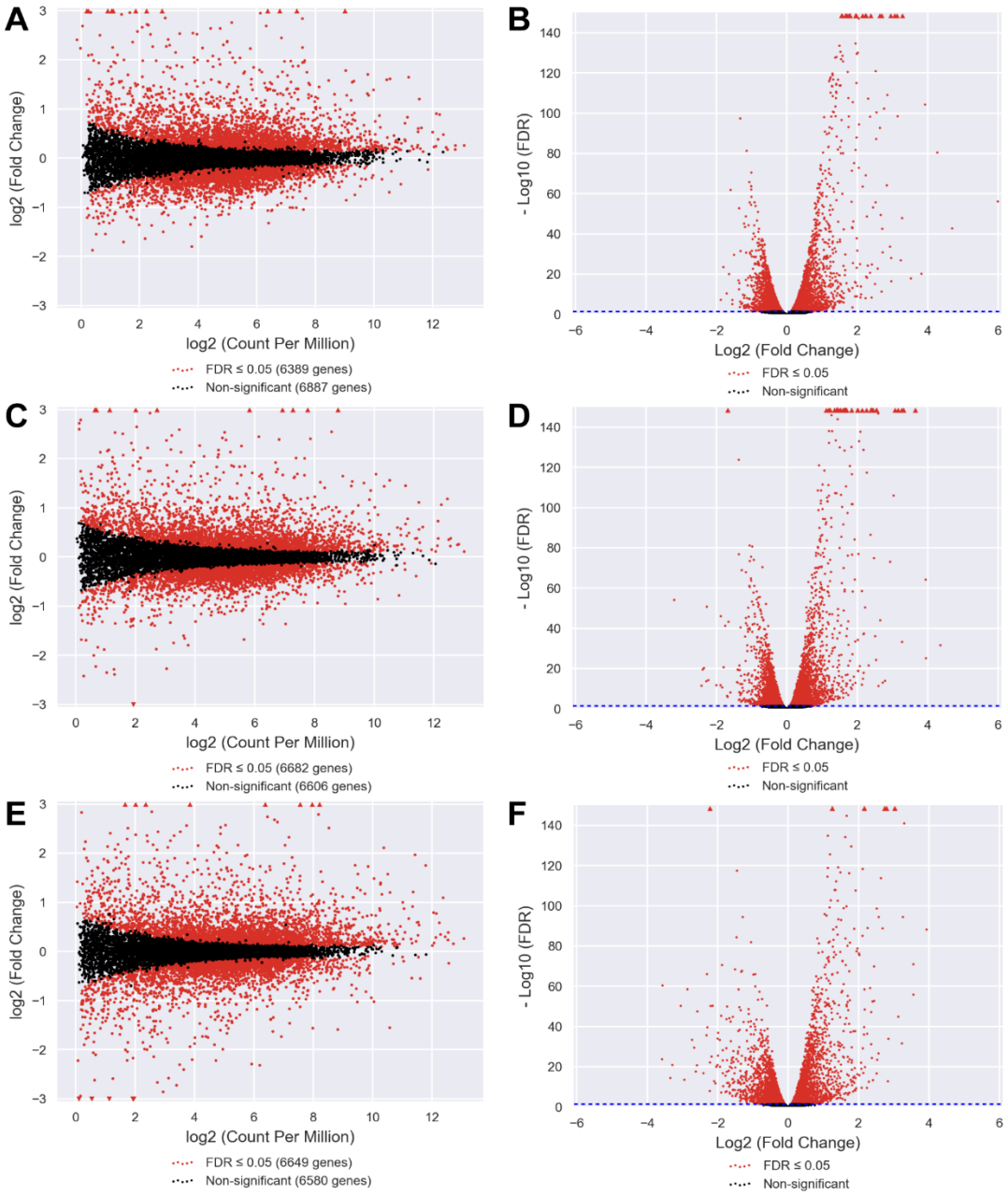

Figure 4.8 *edgeR* comparisons of hypoxia time points and control
EdgeR comparisons between control and 8 hours hypoxia (**A**, **B**), 2 days hypoxia (**C**, **D**) and 2 weeks hypoxia (**E**, **F**). **A.**, **C.**, **E**. Measured CPM and fold-change for each expressed gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. **B.**, **D.**, **F**. *edgeR* measured fold-change and FDR displayed for each gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. Blue dashed line indicates FDR=0.05.

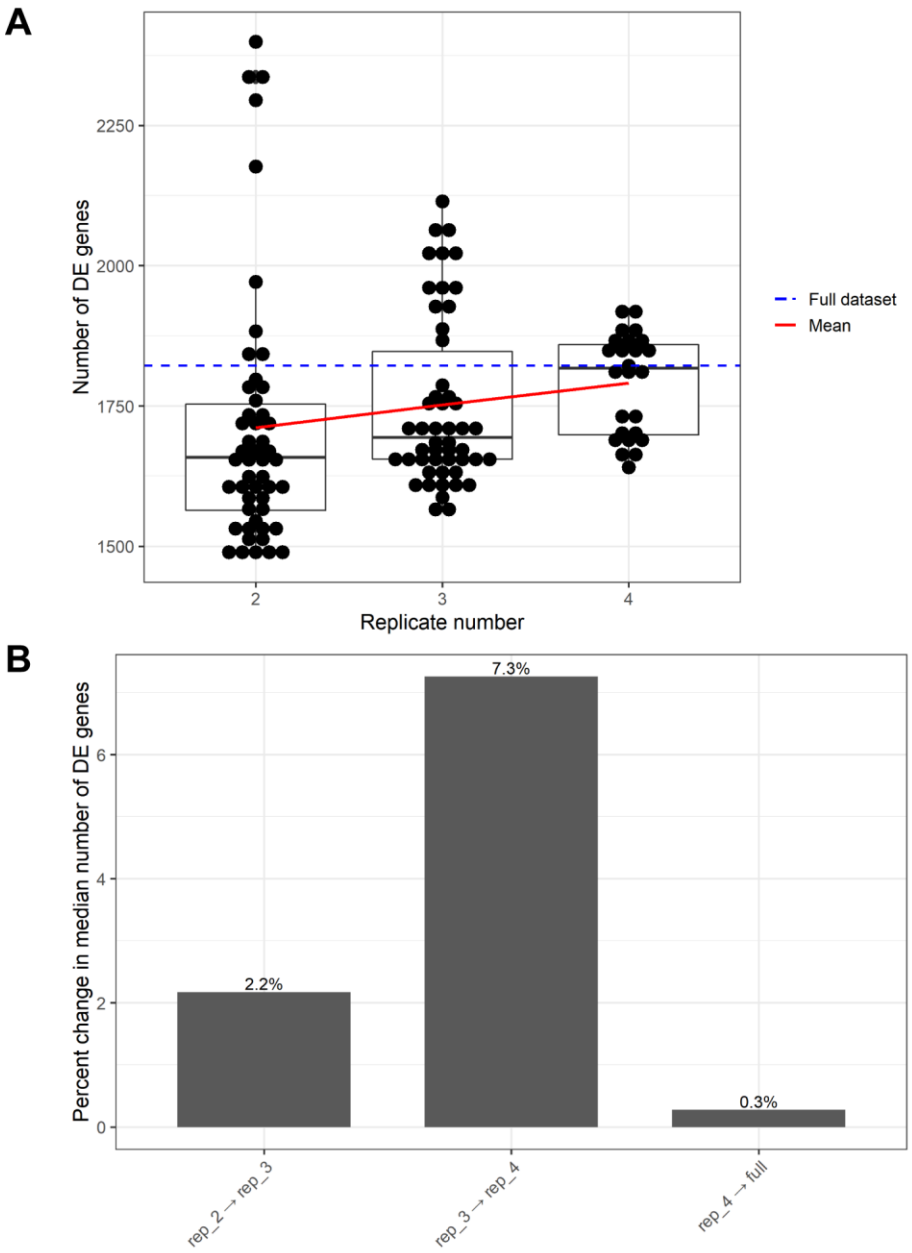

Figure 4.9 *ERSSA* result of the DE comparison between 8 hours hypoxia and control
ERSSA-generated number of DE gene plot (**A**) and marginal change plot (**B**) with 50 combinations per replicate level and log2-fold-change cut off = 0.5. Median used to measure the percentage change in the marginal change plot.

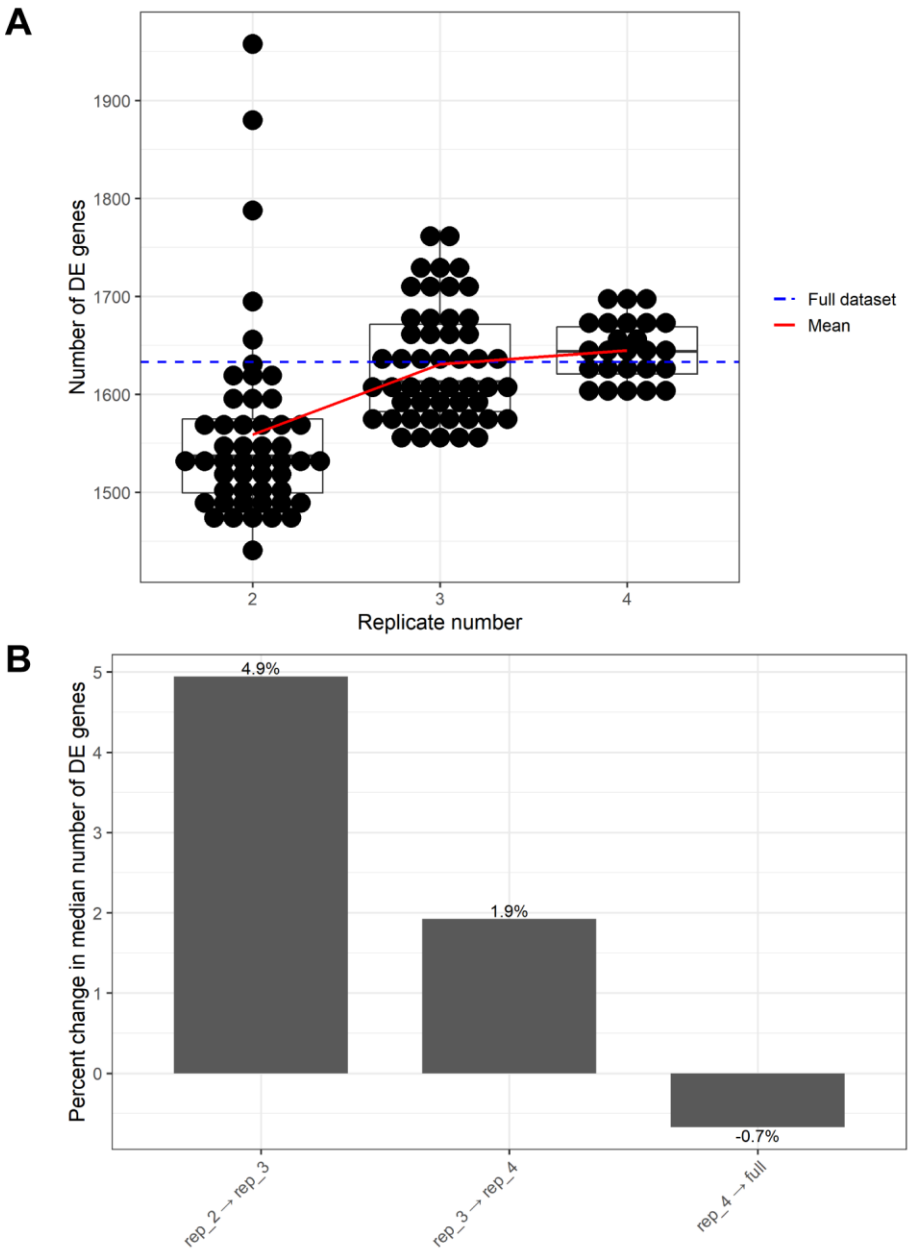

Figure 4.10 *ERSSA* result of the DE comparison between 2 days hypoxia and control
ERSSA-generated number of DE gene plot (**A**) and marginal change plot (**B**) with 50 combinations per replicate level and log2-fold-change cut off = 0.5. Median used to measure the percentage change in the marginal change plot.

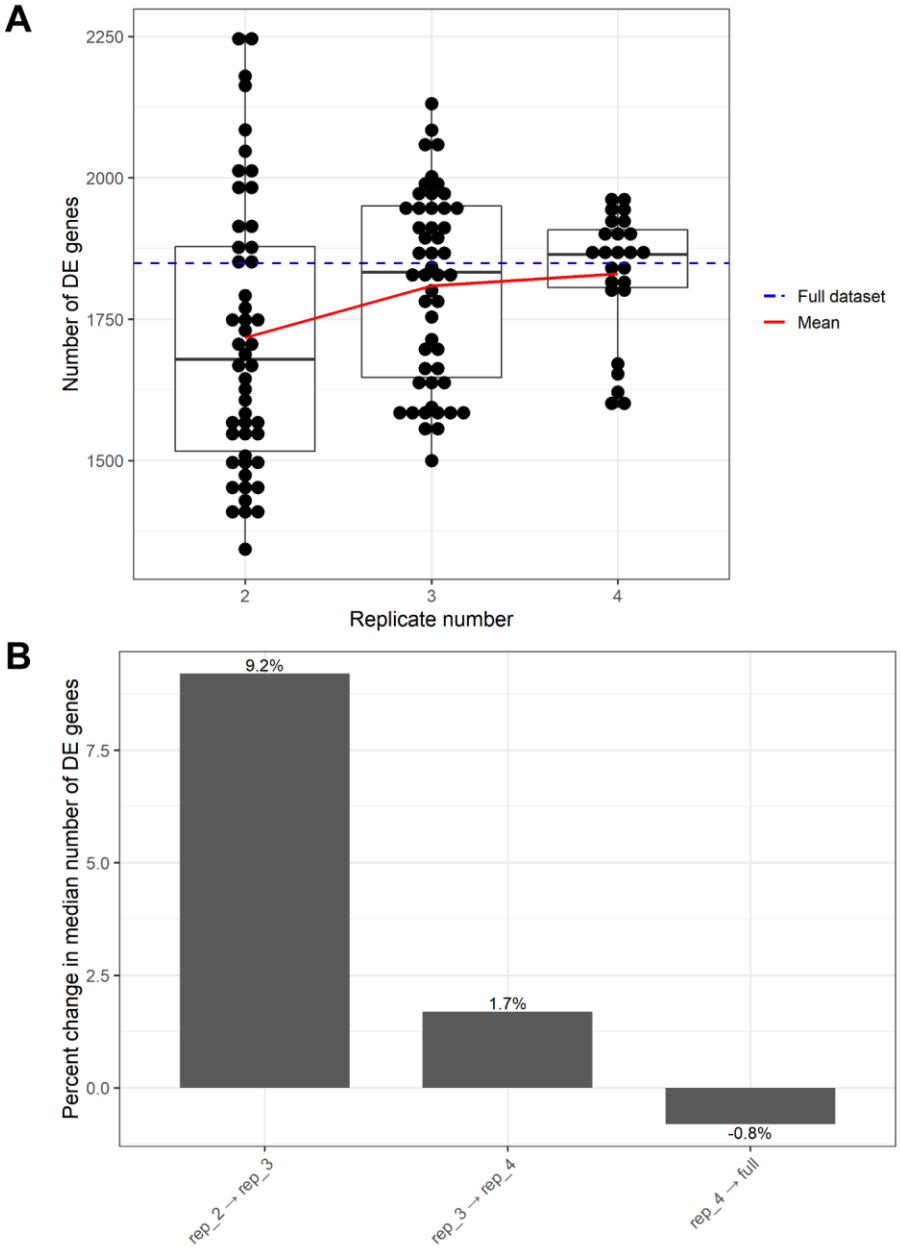

Figure 4.11 *ERSSA* result of the DE comparison between 2 weeks hypoxia and control
ERSSA-generated number of DE gene plot (**A**) and marginal change plot (**B**) with 50 combinations per replicate level and log2-fold-change cut off = 0.5. Median used to measure the percentage change in the marginal change plot.

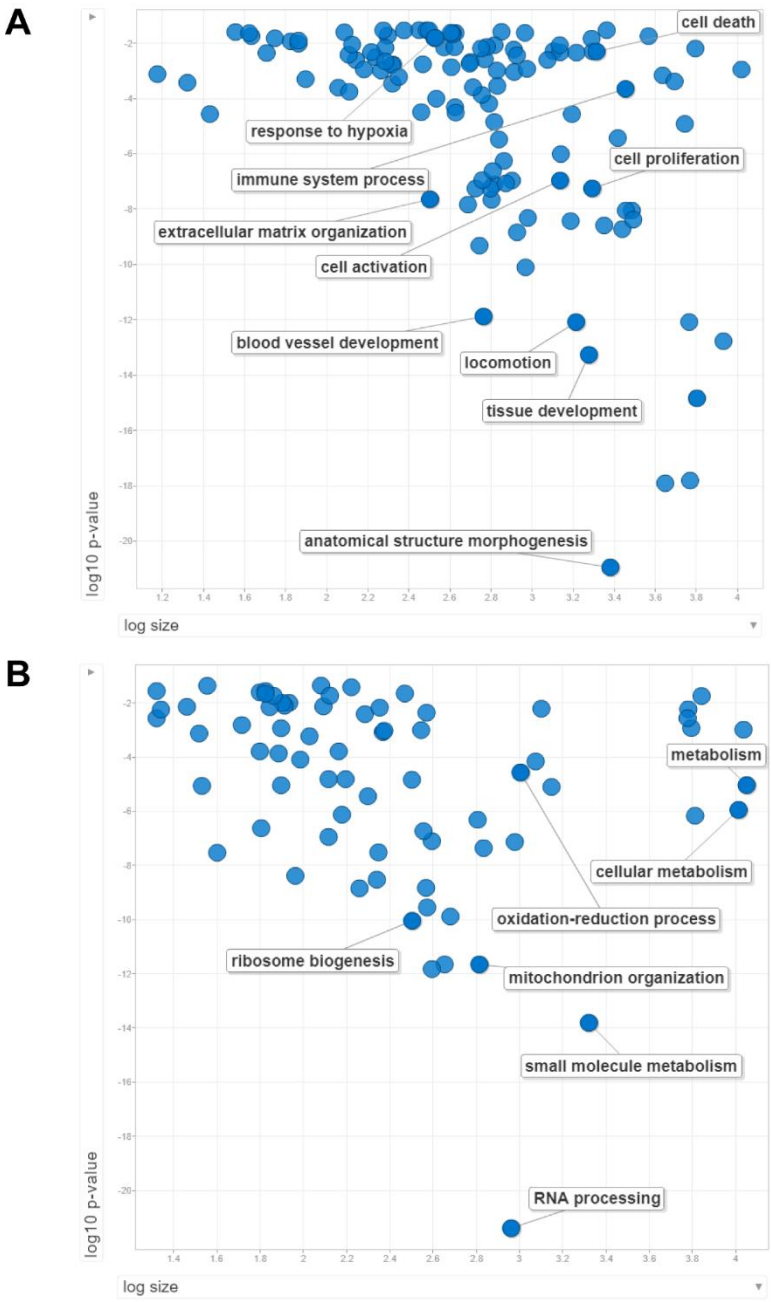

Figure 4.12 GO biological process enrichment with DE genes in comparison between control and 8 hours hypoxia

Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO biological process. Selected biologically relevant GO biological process were labeled.

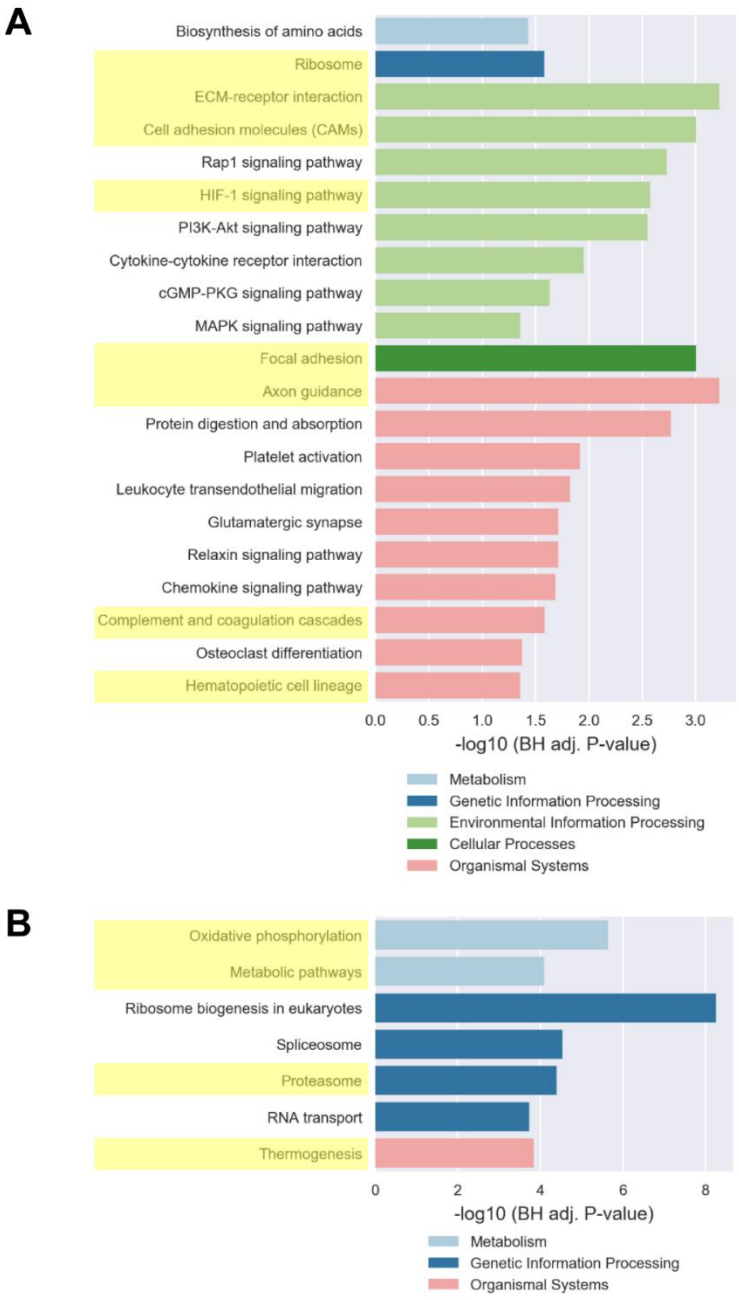

Figure 4.13 KEGG pathway enrichment with DE genes in comparison between control and 8 hours hypoxia

KEGG pathways enriched with *goseq* using up-regulated (**A**) and down-regulated (**B**) DE genes. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis. Pathways enriched at all hypoxia time points are highlighted in yellow.

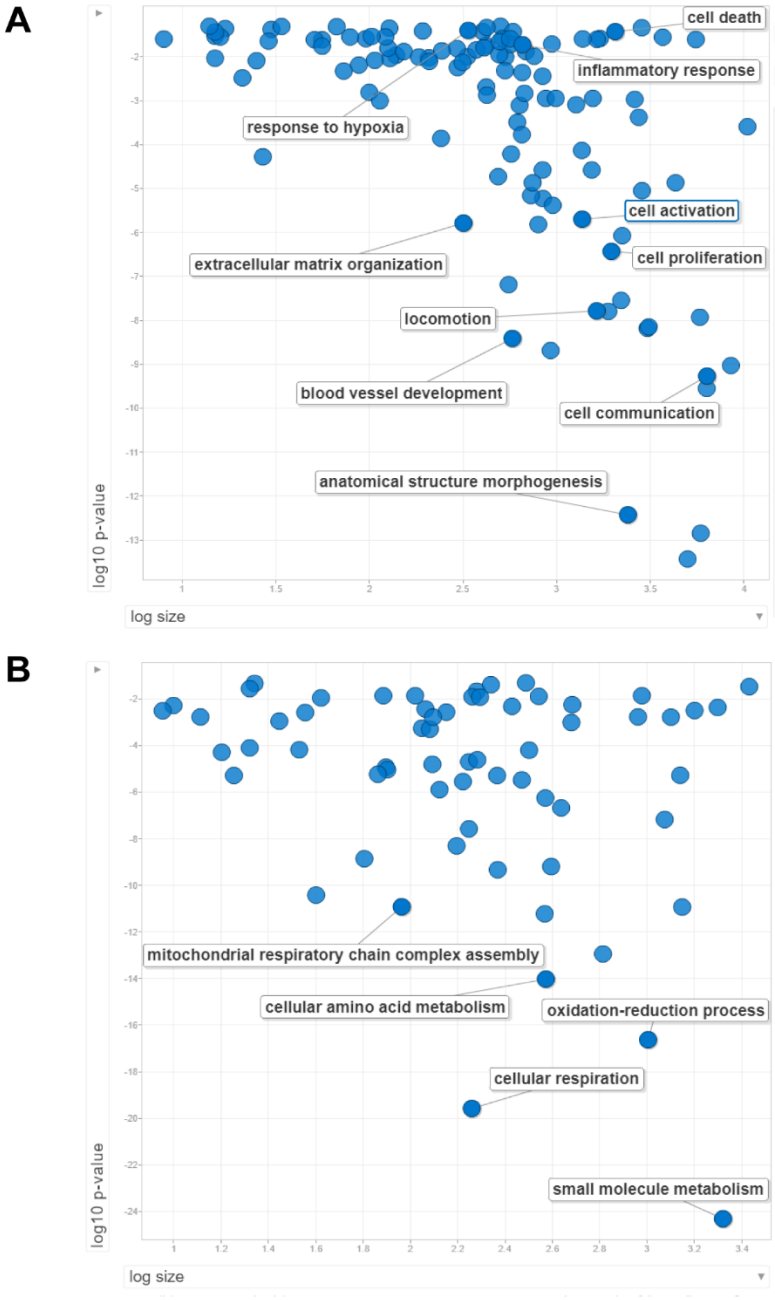

Figure 4.14 GO biological process enrichment with DE genes in comparison between control and 2 days hypoxia

Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO biological process. Selected biologically relevant GO biological process were labeled.

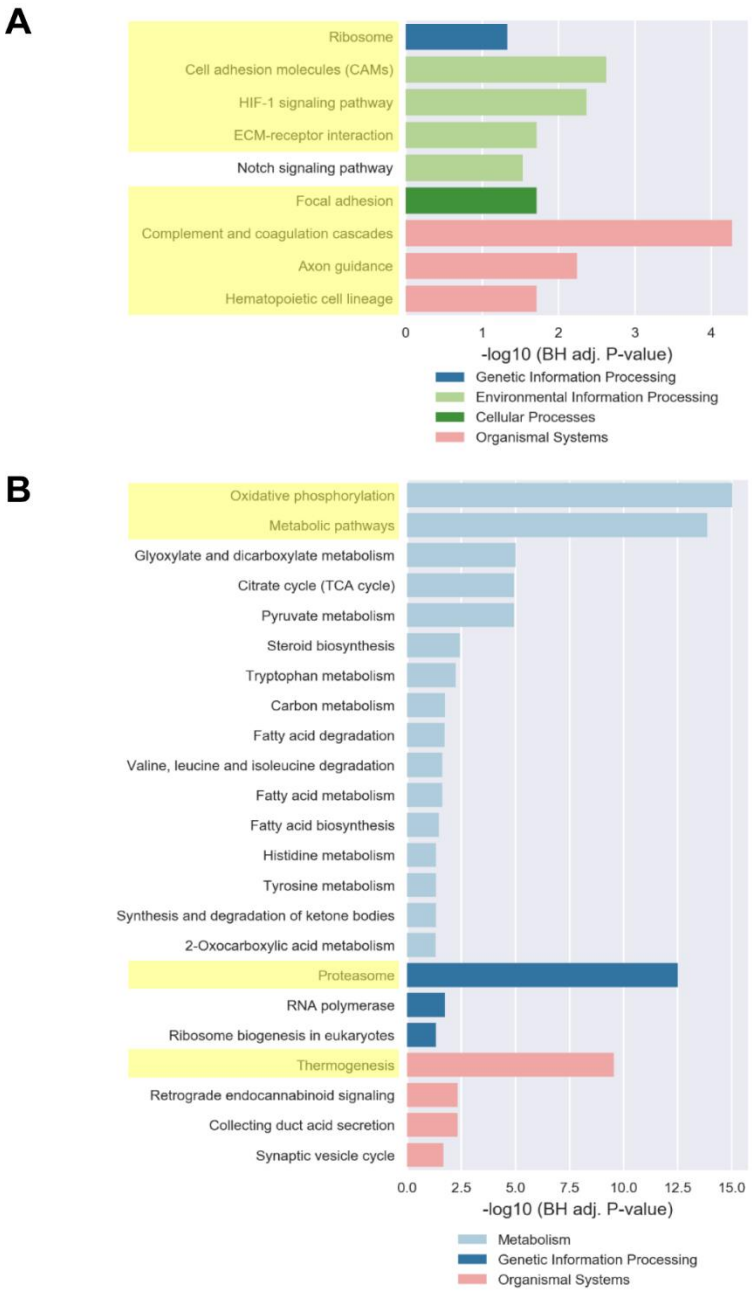

Figure 4.15 KEGG pathway enrichment with DE genes in comparison between control and 2 days hypoxia

KEGG pathways enriched with *goseq* using up-regulated (**A**) and down-regulated (**B**) DE genes. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis. Pathways enriched at all hypoxia time points are highlighted in yellow.

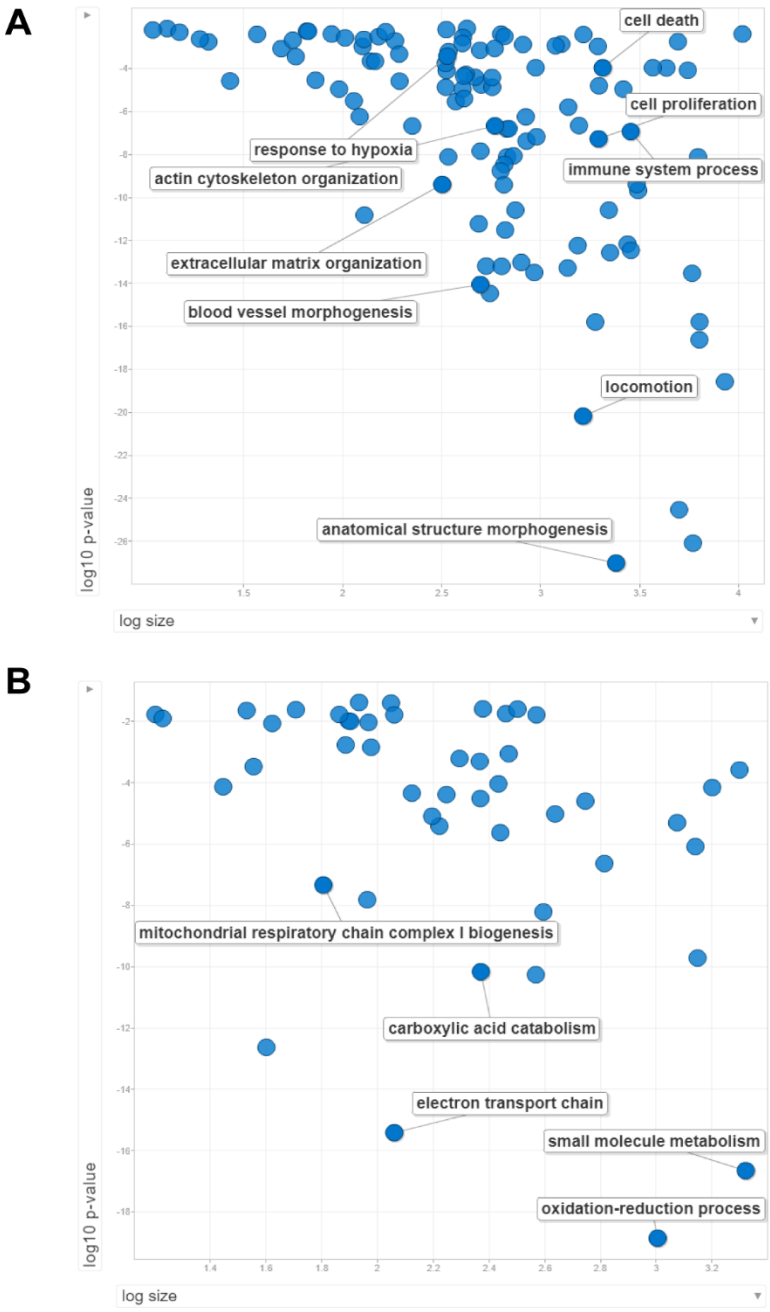

Figure 4.16 GO biological process enrichment with DE genes in comparison between control and 2 weeks hypoxia

Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO biological process. Selected biologically relevant GO biological process were labeled.

Figure 4.17 KEGG pathway enrichment with up-regulated DE genes in comparison between control and 2 weeks hypoxia

KEGG pathways enriched with *goseq* using up-regulated DE genes. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis. Pathways enriched at all hypoxia time points are highlighted in yellow.

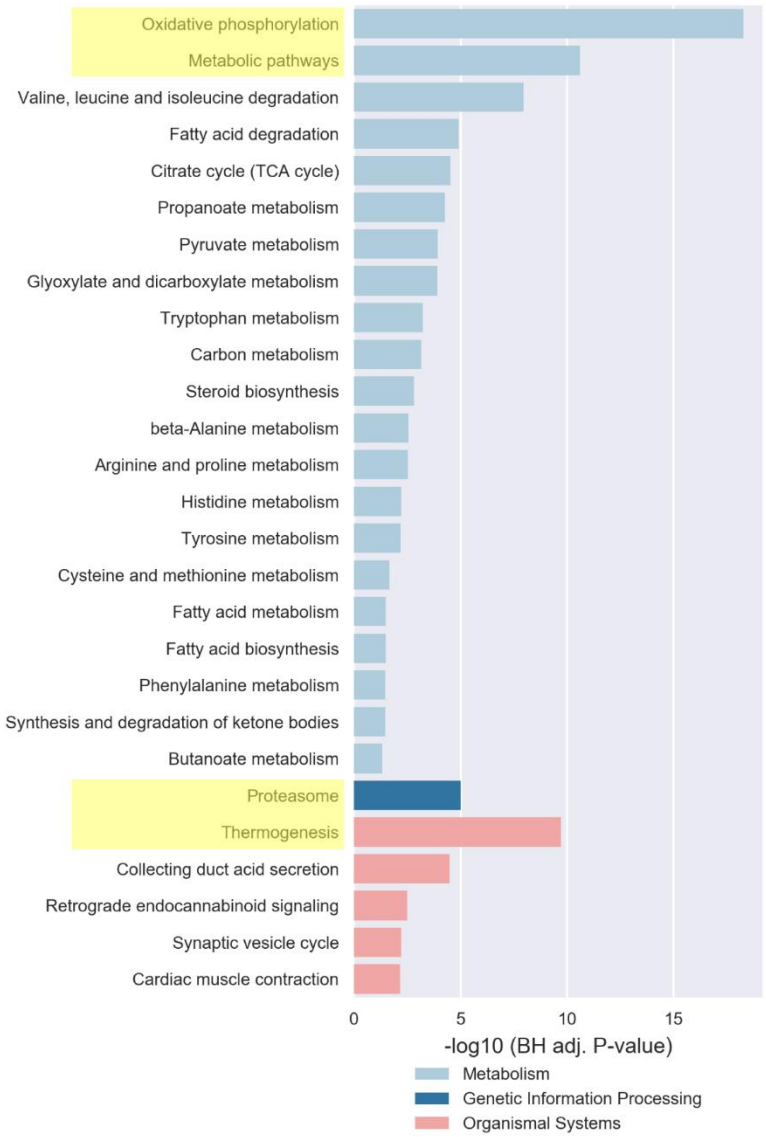

Figure 4.18 KEGG pathway enrichment with down-regulated DE genes in comparison between control and 2 weeks hypoxia

KEGG pathways enriched with *goseq* using down-regulated DE genes. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis. Pathways enriched at all hypoxia time points are highlighted in yellow.

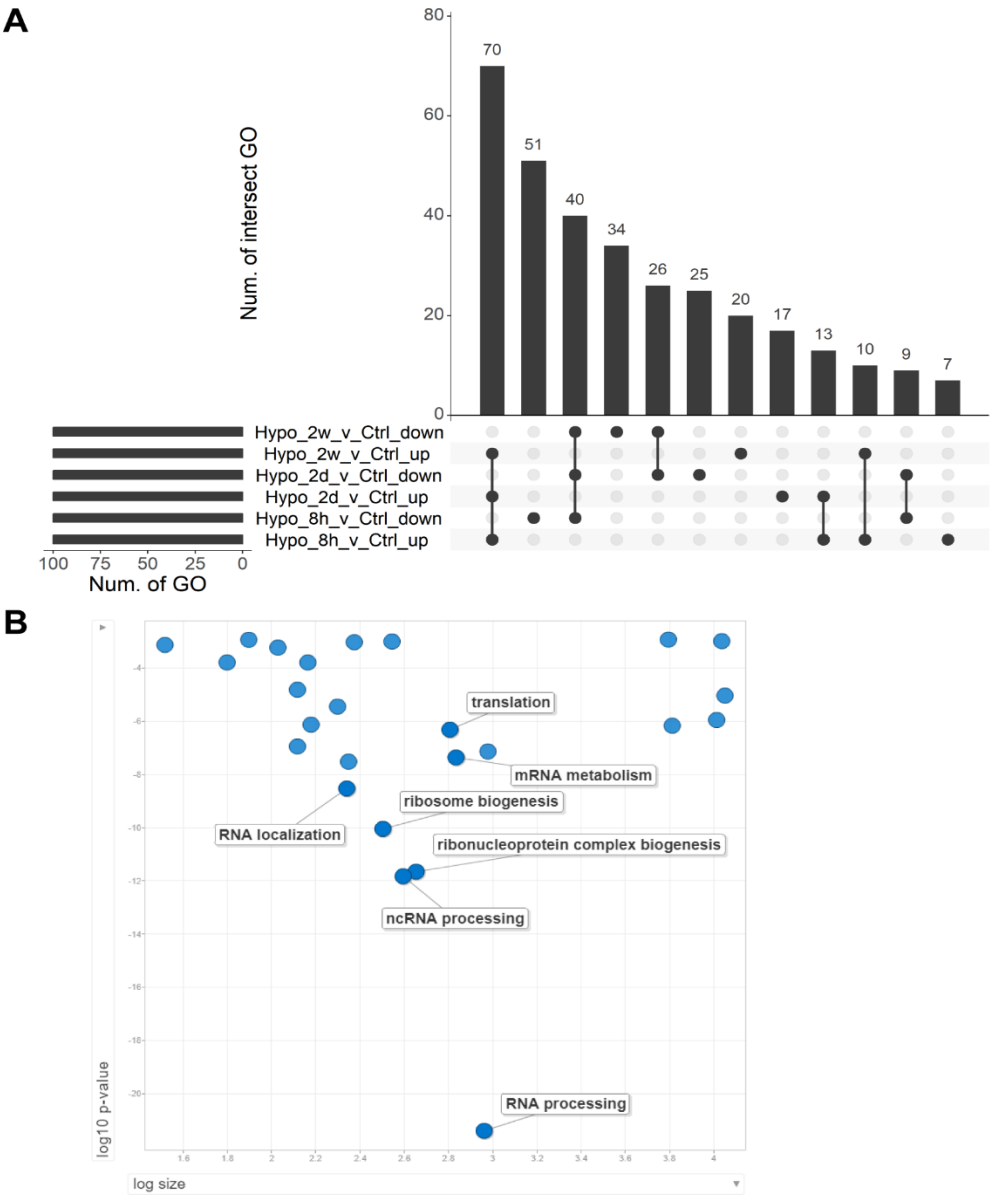

Figure 4.19 Top 100 enriched GO biological processes compared between three comparisons – part 1

Top 100 enriched GO biological processes from each GO enrichment were compared. **A.** Plot shows the overlap between the top 100 lists. Bars indicate the number of overlap among the 6 lists of GO annotations. Dot with connected line indicate inclusion in the GO list. GO lists are named with syntax: Hypoxia_TimePoint_v_Control_Direction with TimePoint of 2w as 2 weeks, 2d as 2 days, and 8h as 8 hours. Direction is either up- or down-regulation. **B.** *REVIGO* visualization of 51 unique GO biological process in the top 100 enriched with 8 hours hypoxia down-regulated genes. Selected biologically relevant processes were labeled.

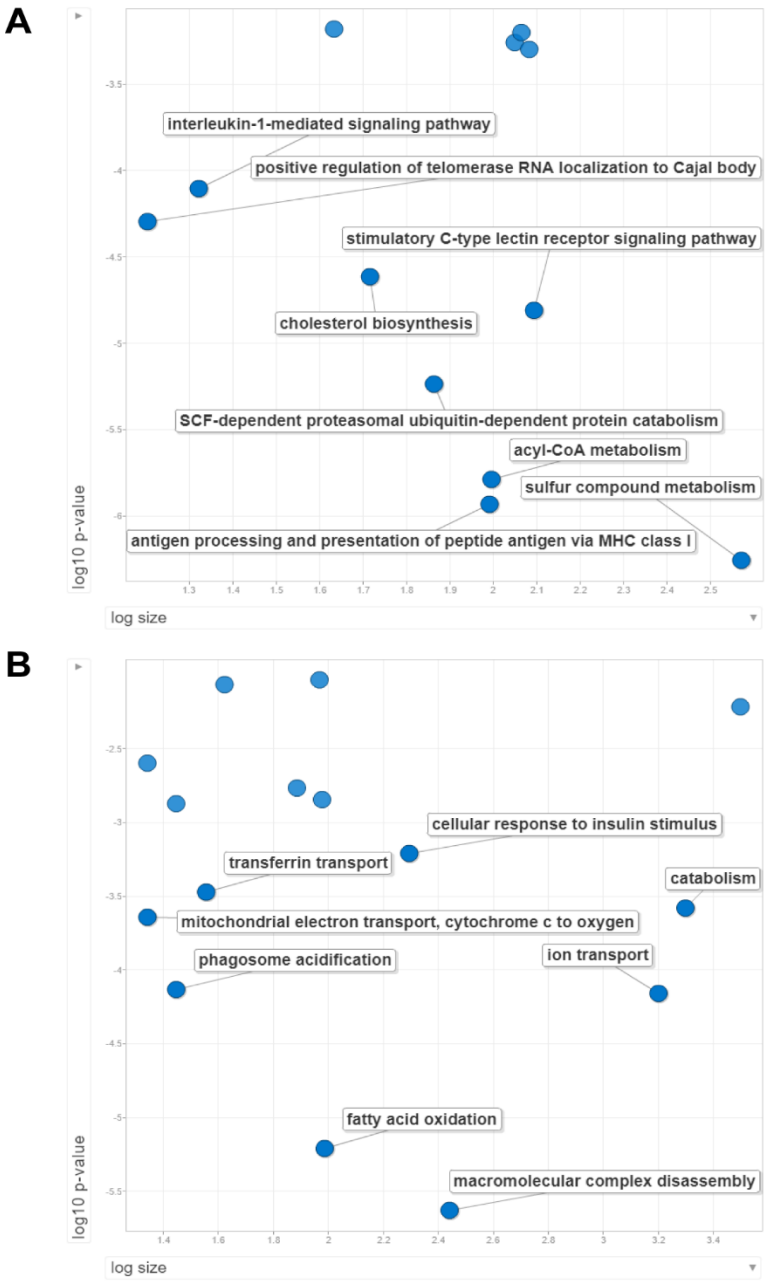

Figure 4.20 Top 100 enriched GO biological processes compared between three comparisons – part 2

A. *REVIGO* visualization of 25 unique processes in the top 100 enriched with 2 days hypoxia down-regulated genes. Selected biologically relevant processes were labeled. **B.** *REVIGO* visualization of 34 unique processes in the top 100 enriched with 2 weeks hypoxia down-regulated genes. Selected biologically relevant processes were labeled.

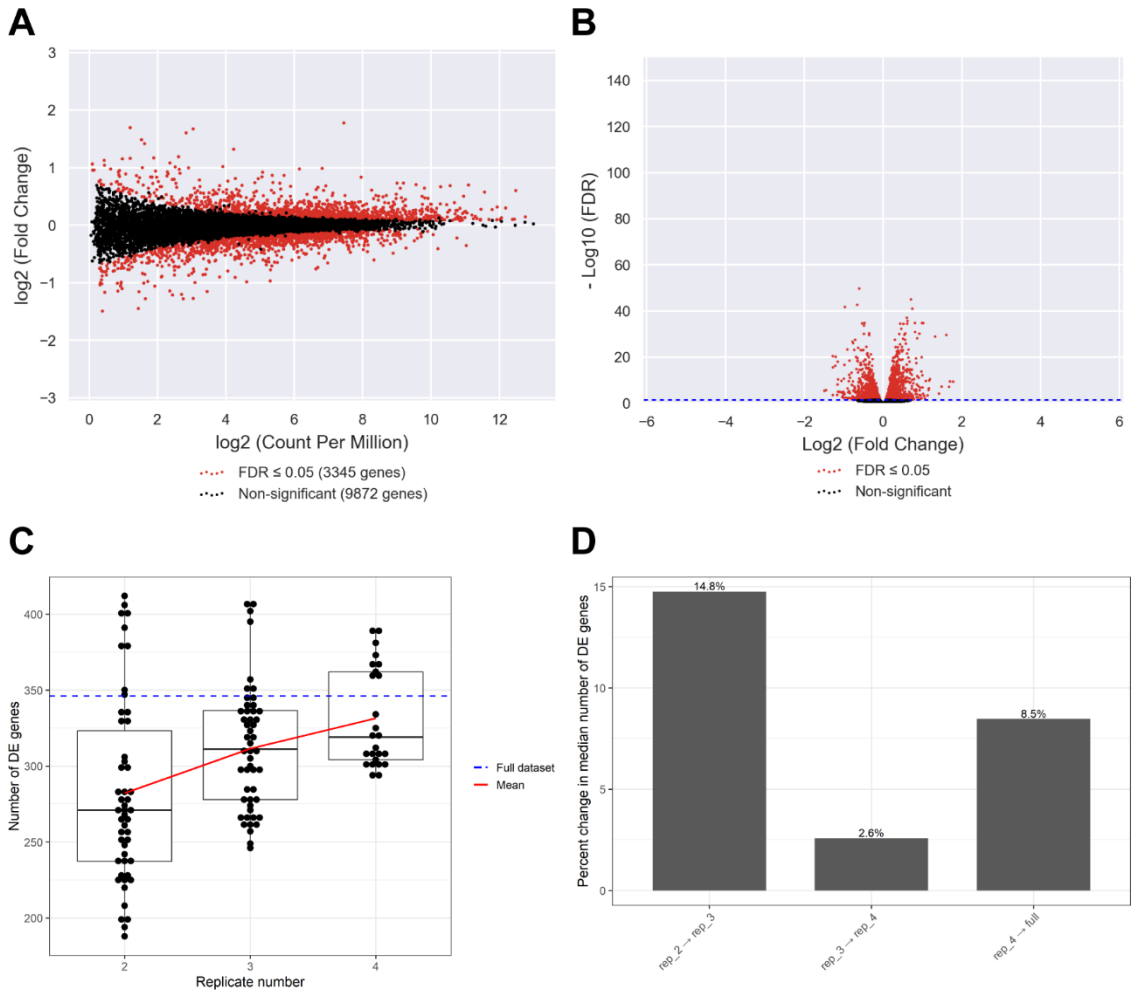

Figure 4.21 *edgeR* and *ERSSA* results of DE comparison between hypoxia recovery and control

A. *edgeR*-measured CPM and fold-change for each expressed gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. **B.** *edgeR*-measured fold-change and FDR displayed for each gene. DE genes are colored red while all others are black. Genes outside of display window are symbolized by triangle. Blue dash line indicates FDR=0.05. **C-D.** *ERSSA*-generated number of DE gene plot (**C**) and marginal change plot (**D**) with 50 combinations per replicate level and log2-fold-change cut off = 0.5. Median used to measure the percentage change in the marginal change plot.

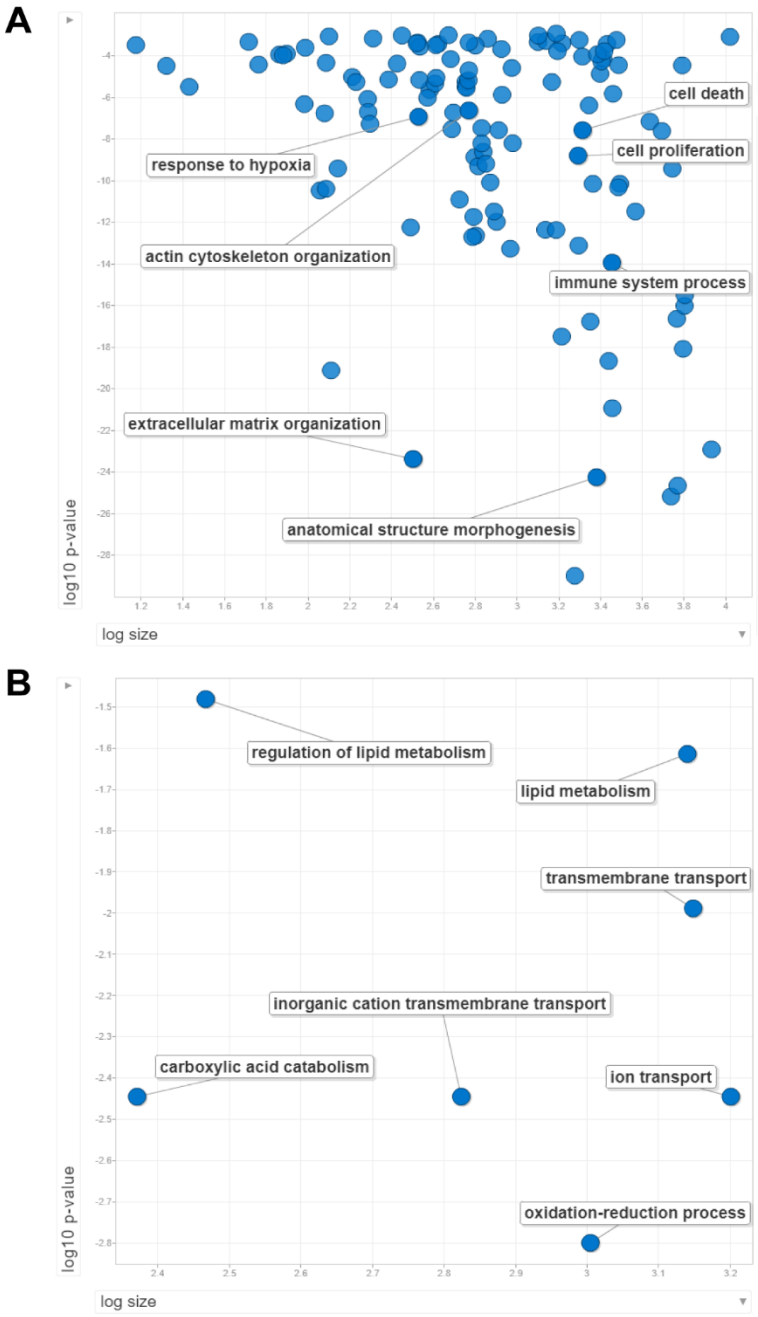

Figure 4.22 GO biological process enrichment with DE genes in comparison between control and hypoxia recovery

Up- (**A**) and down-regulated (**B**) DE genes were submitted for GO biological process enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment p-value. Size is the number of human proteins associated with each GO biological process. Selected biologically relevant processes were labeled.

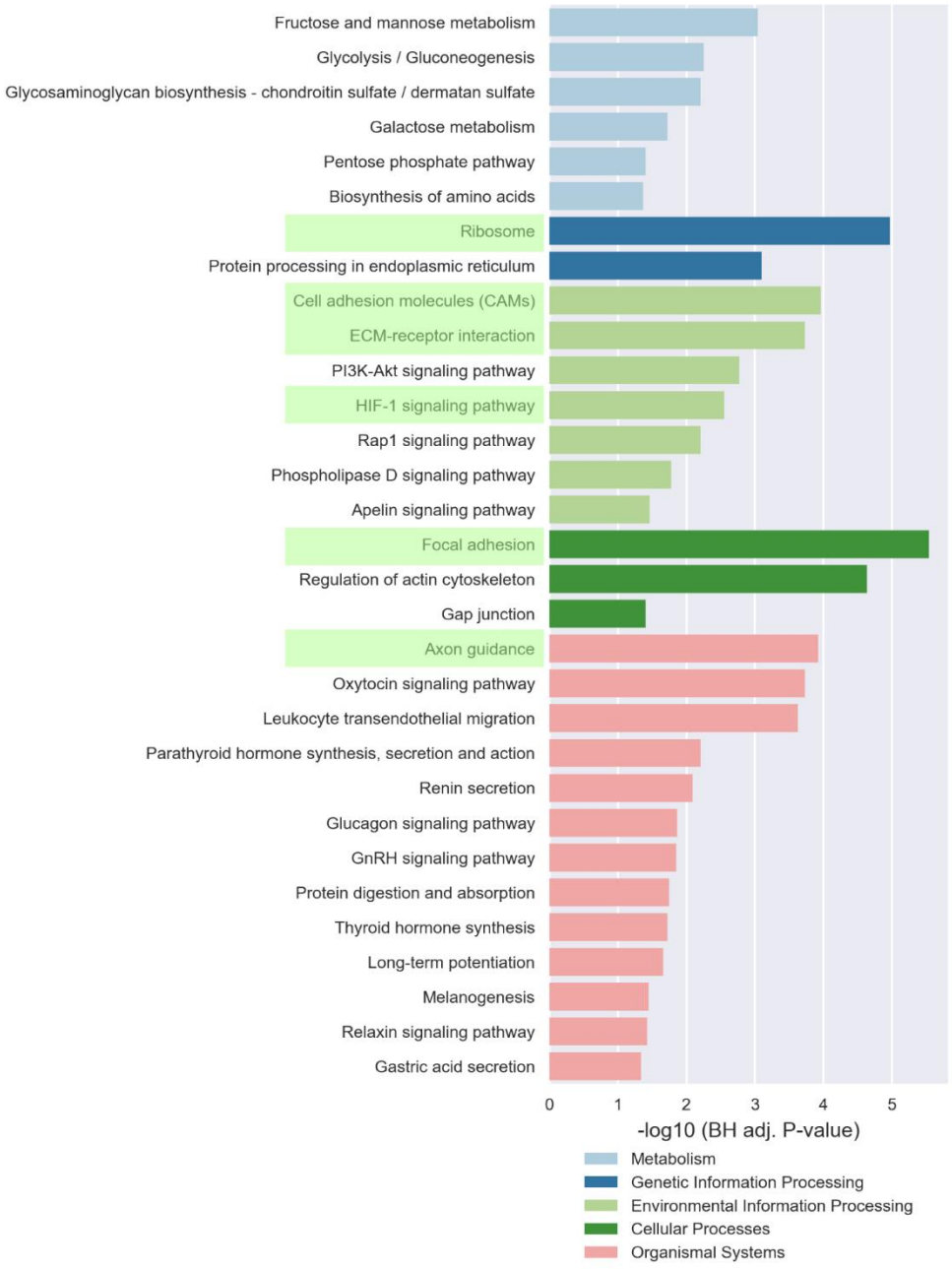

Figure 4.23 KEGG pathway enrichment with up-regulated DE genes in comparison between control and hypoxia recovery

KEGG pathways enriched with *goseq* using up-regulated DE genes. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis. Pathways also enriched at all hypoxia time points without recovery are highlighted in green.

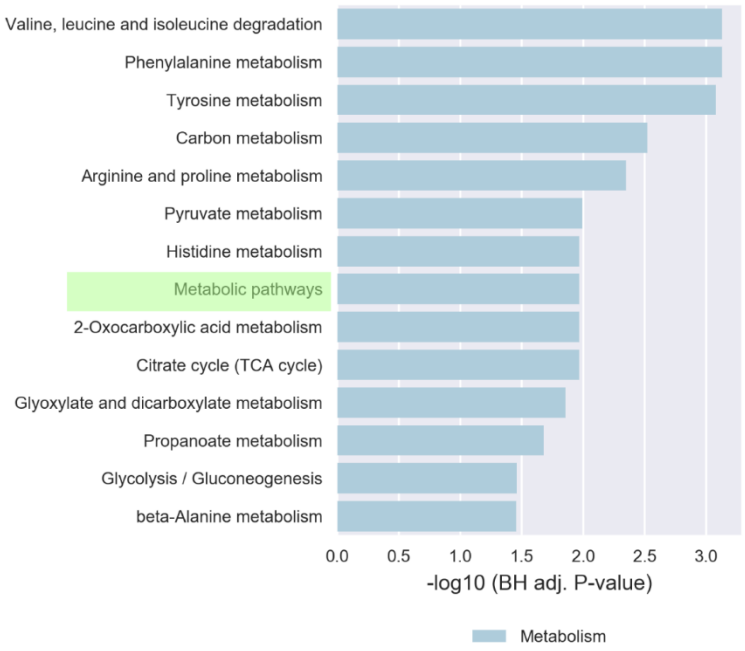

Figure 4.24 KEGG pathway enrichment with down-regulated DE genes in comparison between control and hypoxia recovery

KEGG pathways enriched with *goseq* using down-regulated DE genes. Statistically significant pathways are displayed. Benjamini Hochberg (BH) adjusted P-value is displayed on the x-axis. Pathways also enriched at all hypoxia time points without recovery are highlighted in green.

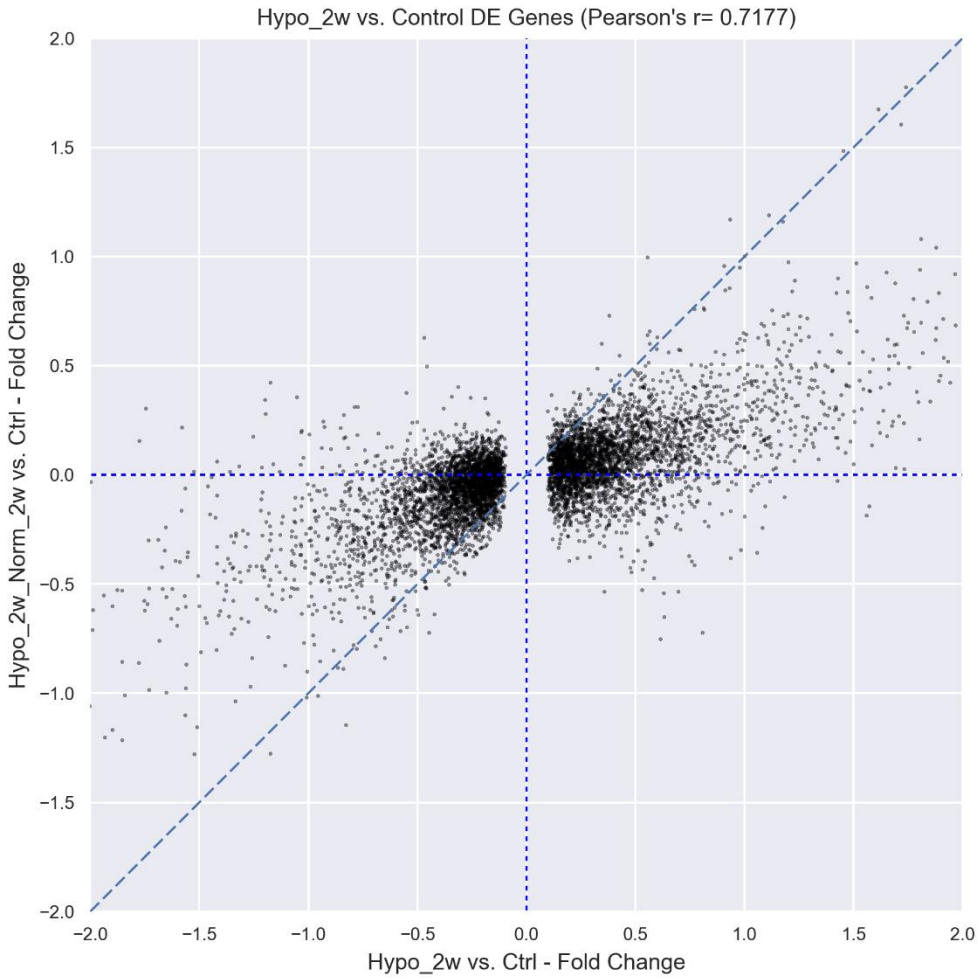

Figure 4.25 Fold change of 2 weeks hypoxia-regulated DE gene after hypoxia recovery
Each dot represent a DE gene found to be regulated after 2 weeks hypoxia treatment. The gene's log2-fold-change after 2 weeks hypoxia treatment is plotted on the x-axis, while the fold change after hypoxia recovery is plotted on the y-axis. The plot window is limited to the (-2, -2) range on both axes. Pearson's r value is measured as 0.7177. It is worth noting that a majority of genes sits below y=x line (dash dark blue line) in positive x-axis regime and above y=x line in negative x-axis regime, suggesting hypoxia recovery moderated hypoxia-induced transcriptome regulation.

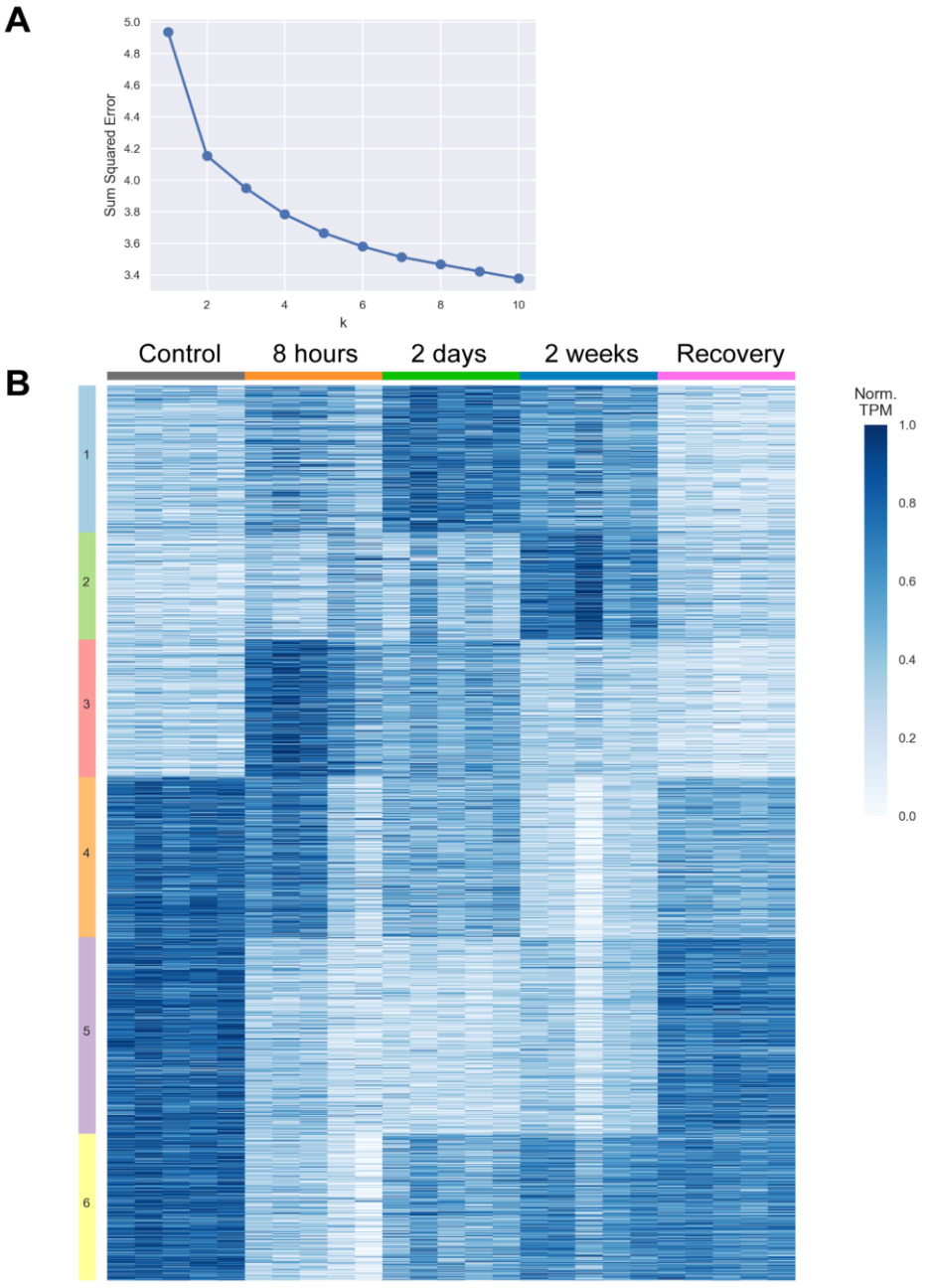

Figure 4.26 Expression profile of all DE genes clustered by k-mean method
A. Sum squared error of k-mean clustering with up to k=10 clusters. k=6 selected for further cluster-specific analysis. **B.** Gene-wise normalized (Norm.) TPM plotted as a heatmap with genes on rows and samples organized by conditions (color-coded on top of heatmap) on columns. Genes sorted by k-mean clustering result (k=6). Cluster association indicated by color and number on the left of heatmap.

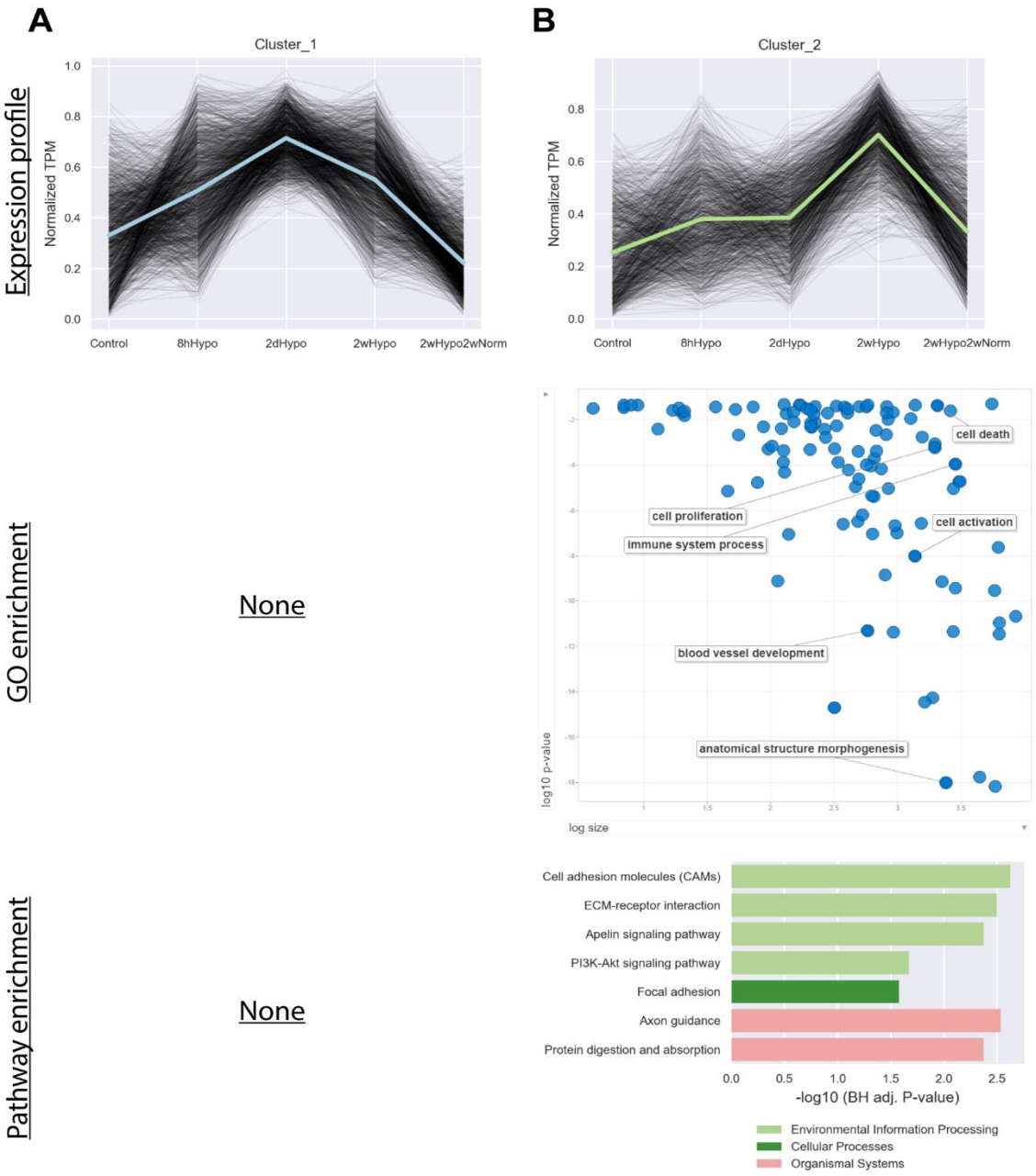

Figure 4.27 Cluster-specific expression profile and enrichment
For cluster 1 (**A**) or cluster 2 (**B**), every gene's condition-averaged expression profile is plotted in black and the average across genes plotted as a thick colored line. Cluster associated genes are used for GO biological process enrichment with results visualized with *REVIGO*. Statistically significant pathways enriched are also displayed.

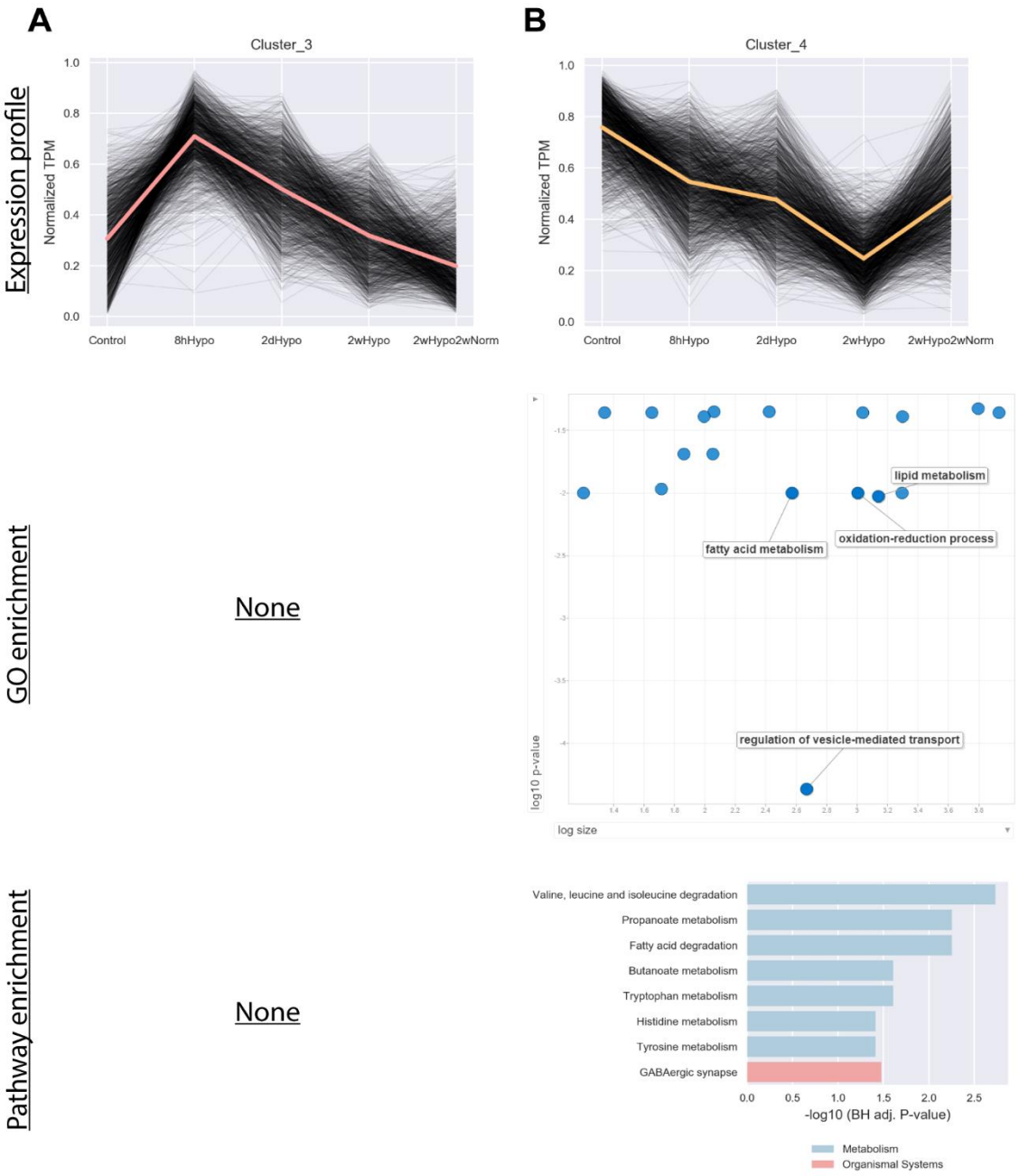

Figure 4.28 Cluster-specific expression profile and enrichment
For cluster 3 (**A**) or cluster 4 (**B**), every gene's condition-averaged expression profile is plotted in black and the average across genes plotted as a thick colored line. Cluster associated genes are used for GO biological process enrichment with results visualized with *REVIGO*. Statistically significant pathways enriched are also displayed.

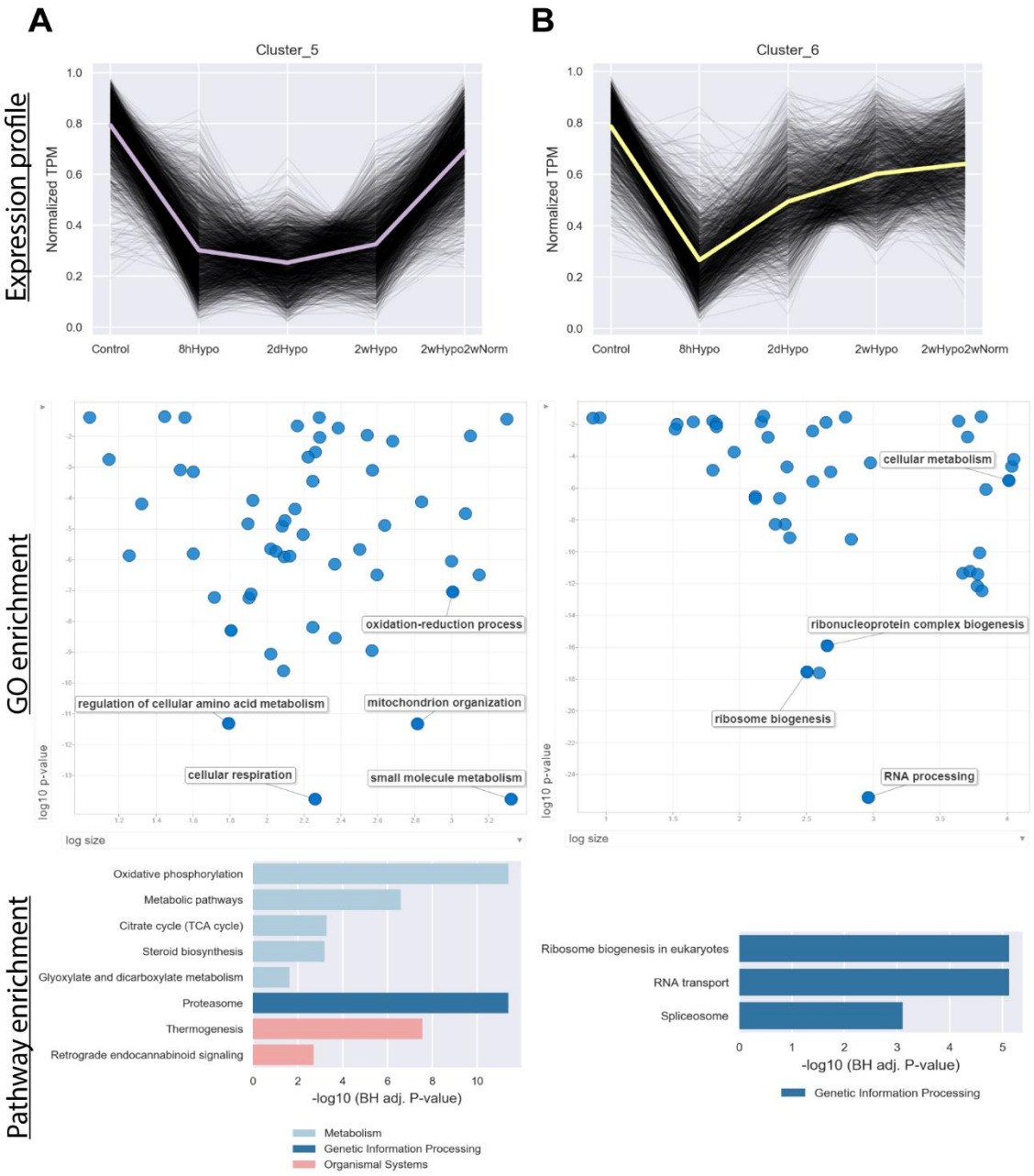

Figure 4.29 Cluster-specific expression profile and enrichment
For cluster 5 (**A**) or cluster 6 (**B**), every gene's condition-averaged expression profile is plotted in black and the average across genes plotted as a thick colored line. Cluster associated genes are used for GO biological process enrichment with results visualized with *REVIGO*. Statistically significant pathways enriched are also displayed.

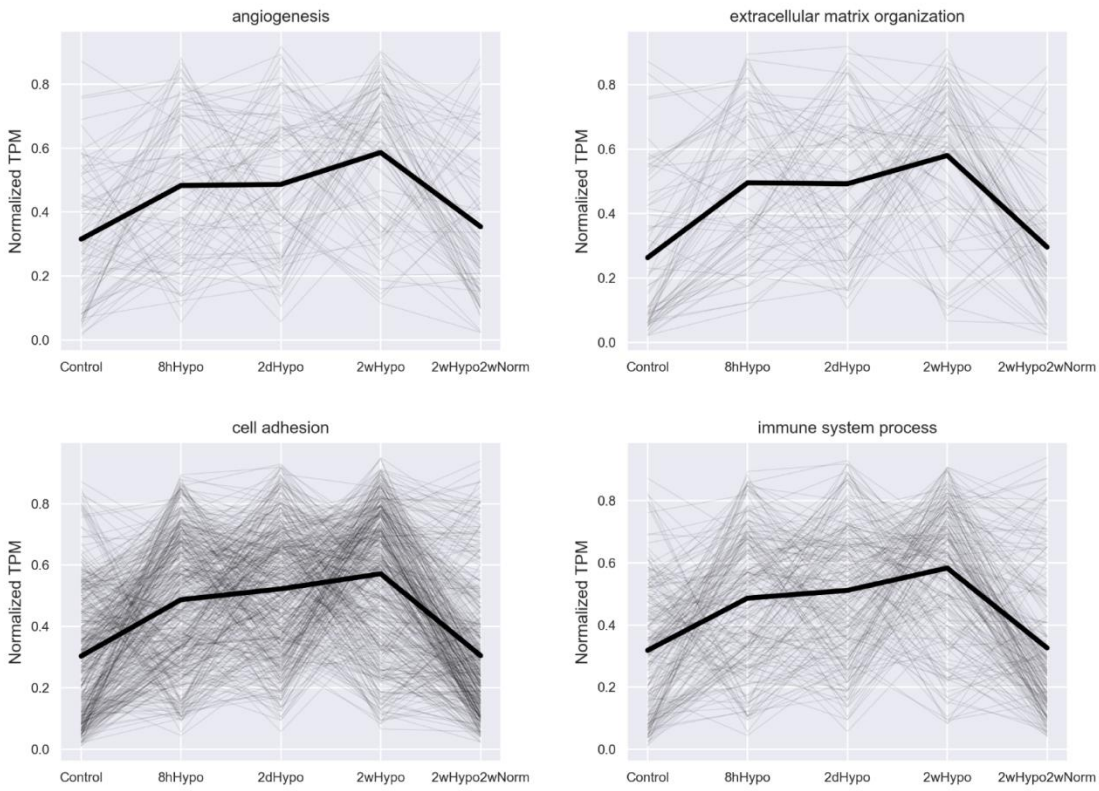

Figure 4.30 Expression profile across selected GO

All DE genes in cluster 1, 2 and 3 that belong to one of the four GO biological processes selected were used to plot individual gene expression profile in gray lines and average expression in thick solid black line. For all four GO biological processes, the average selected genes' expression is elevated after 8 hours and 2 days hypoxia and reached peak expression after 2 weeks hypoxia. 8hHypo is 8 hours hypoxia, 2dHypo is 2 days hypoxia, 2wHypo is 2 weeks hypoxia, and 2wHypo2wNorm is hypoxia recovery conditions.

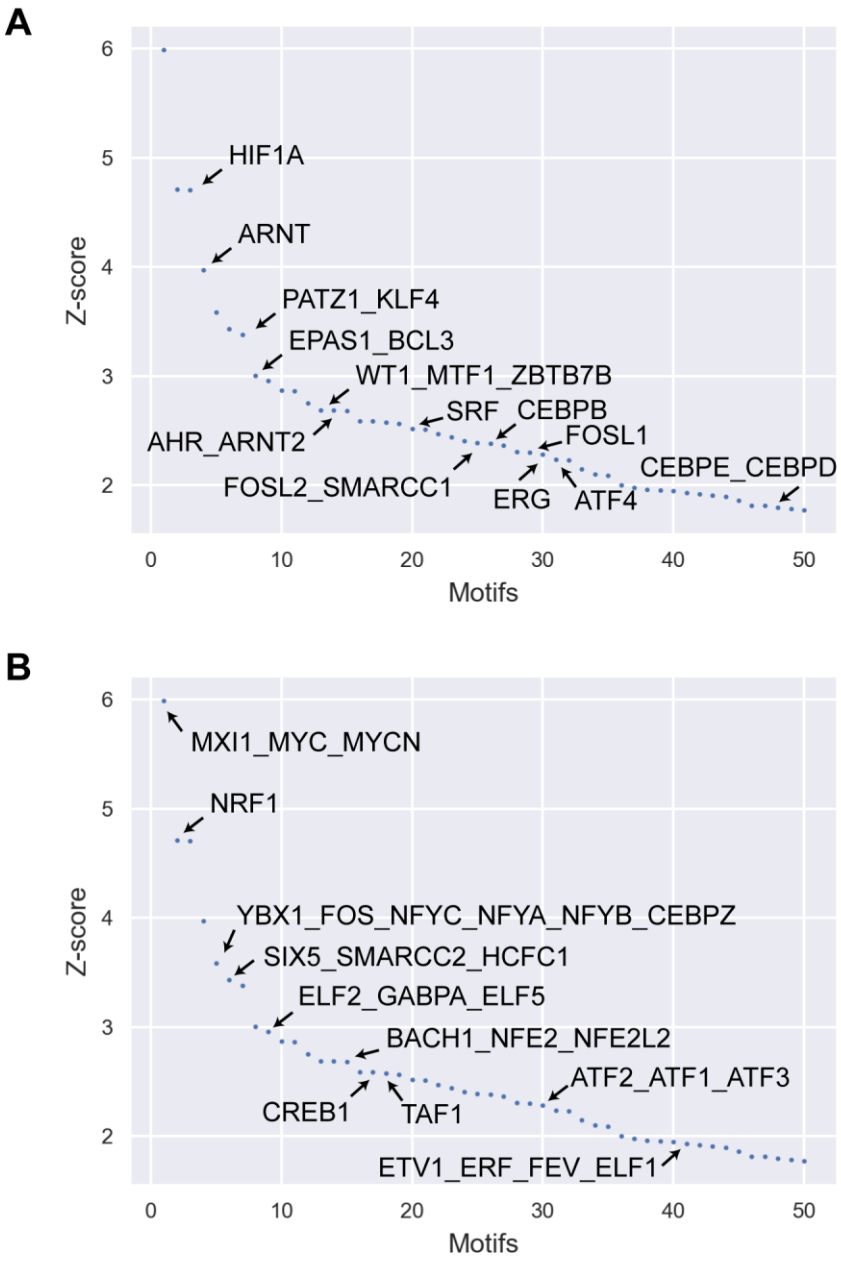

Figure 4.31 *ISMARA* TF enrichment
All samples were submitted to *ISMARA* for enrichment of TFs that are involved in ARPE-19 hypoxia response. Top 50 enriched factors as ranked by Z-score are plotted (x-axis shows rank by Z-score). Selected biological relevant factors with higher (**A**) or lower (**B**) target gene activity in hypoxia response are labeled.

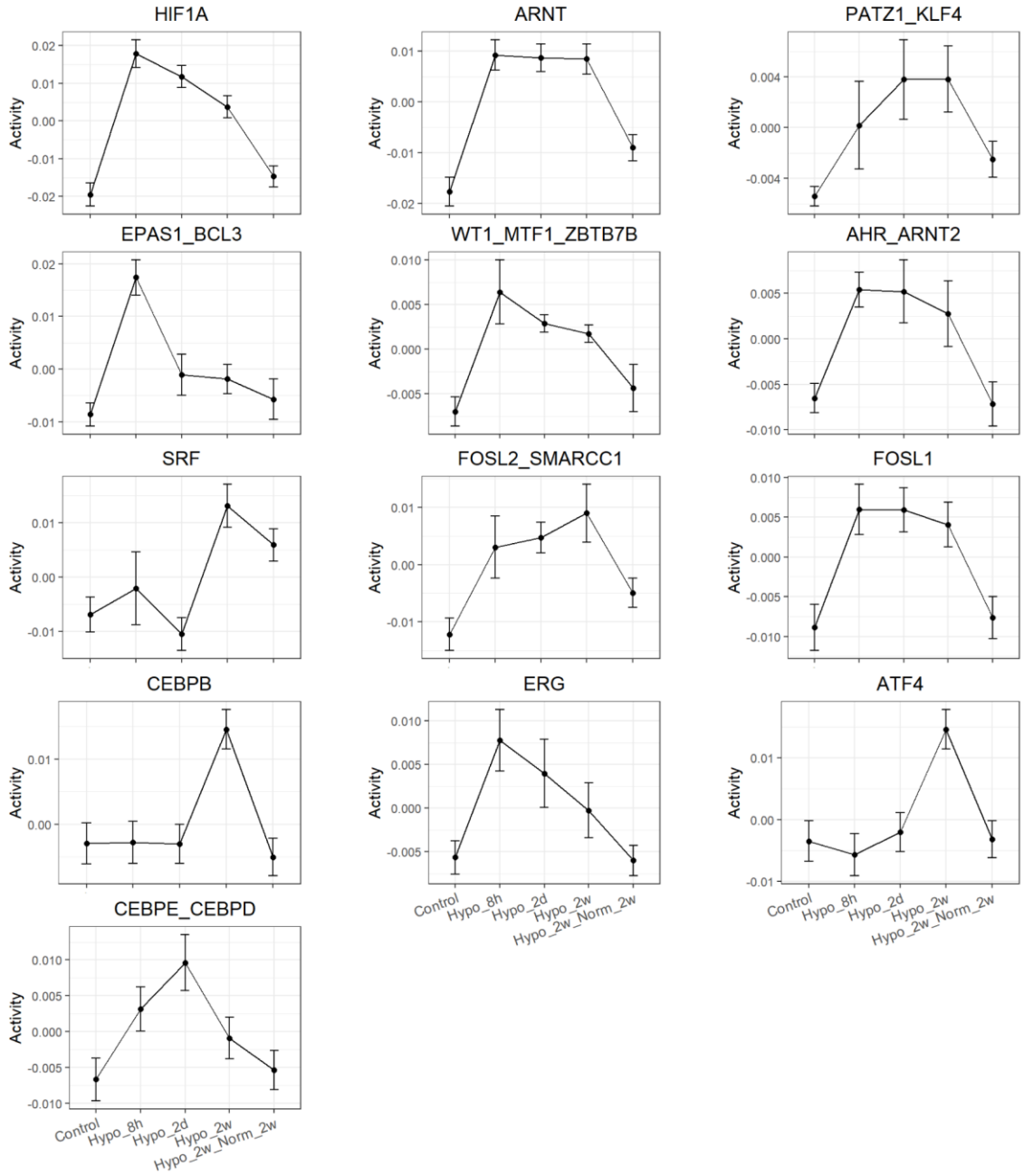

Figure 4.32 Activity profiles of selected TF motifs with higher target gene activity during hypoxia

Activity profile (mean + standard deviation) of selected TFs with higher target gene activity are plotted. Hypo_8h is 8 hours hypoxia, Hypo_2d is 2 days hypoxia, Hypo_2w is 2 weeks hypoxia, and Hypo_2w_Norm_2w is hypoxia recovery conditions.

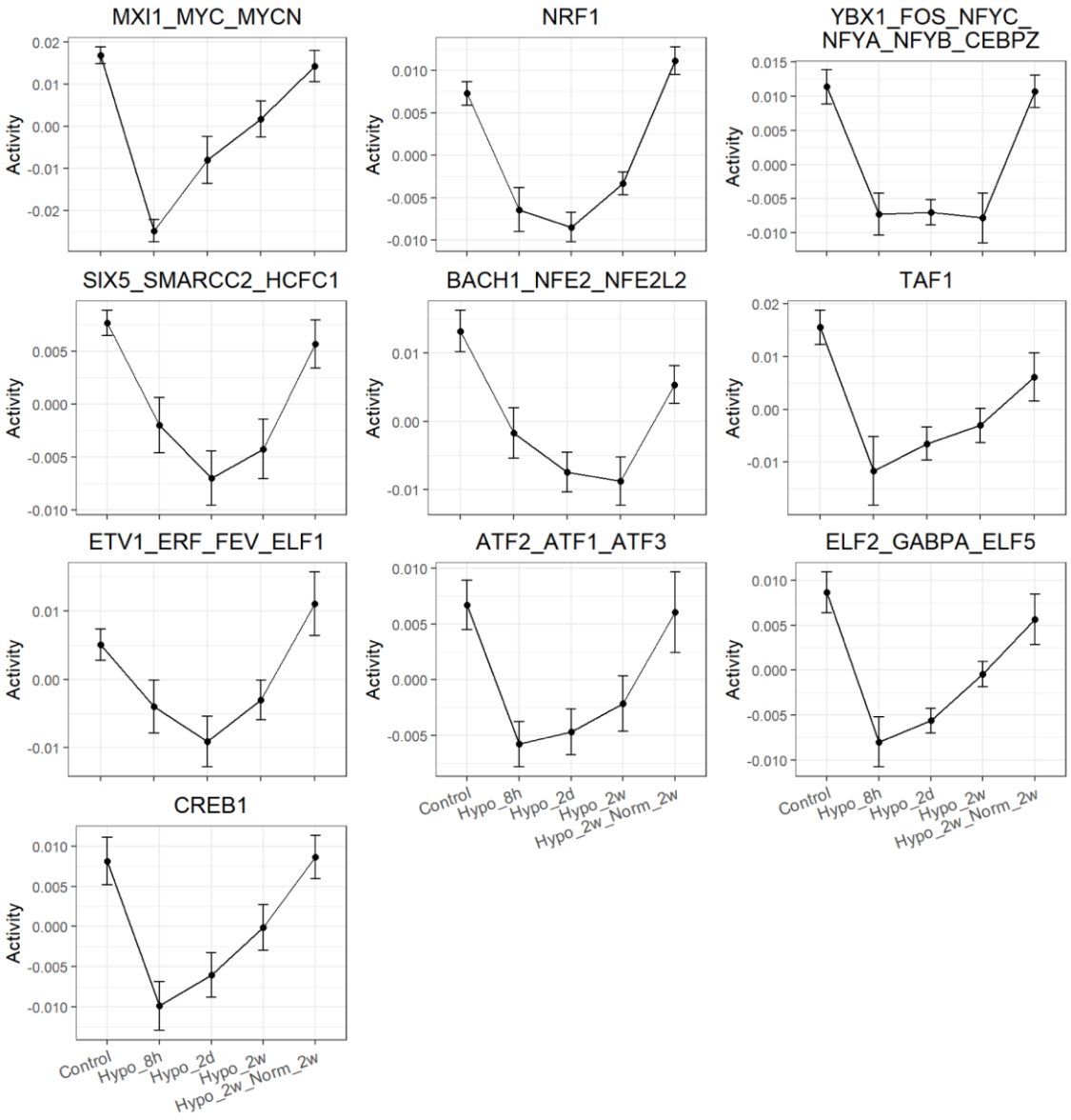

Figure 4.33 Activity profiles of selected TF motifs with lower target gene activity during hypoxia

Activity profile (mean + standard deviation) of selected TFs with lower target gene activity are plotted. Hypo_8h is 8 hours hypoxia, Hypo_2d is 2 days hypoxia, Hypo_2w is 2 weeks hypoxia, and Hypo_2w_Norm_2w is hypoxia recovery conditions.

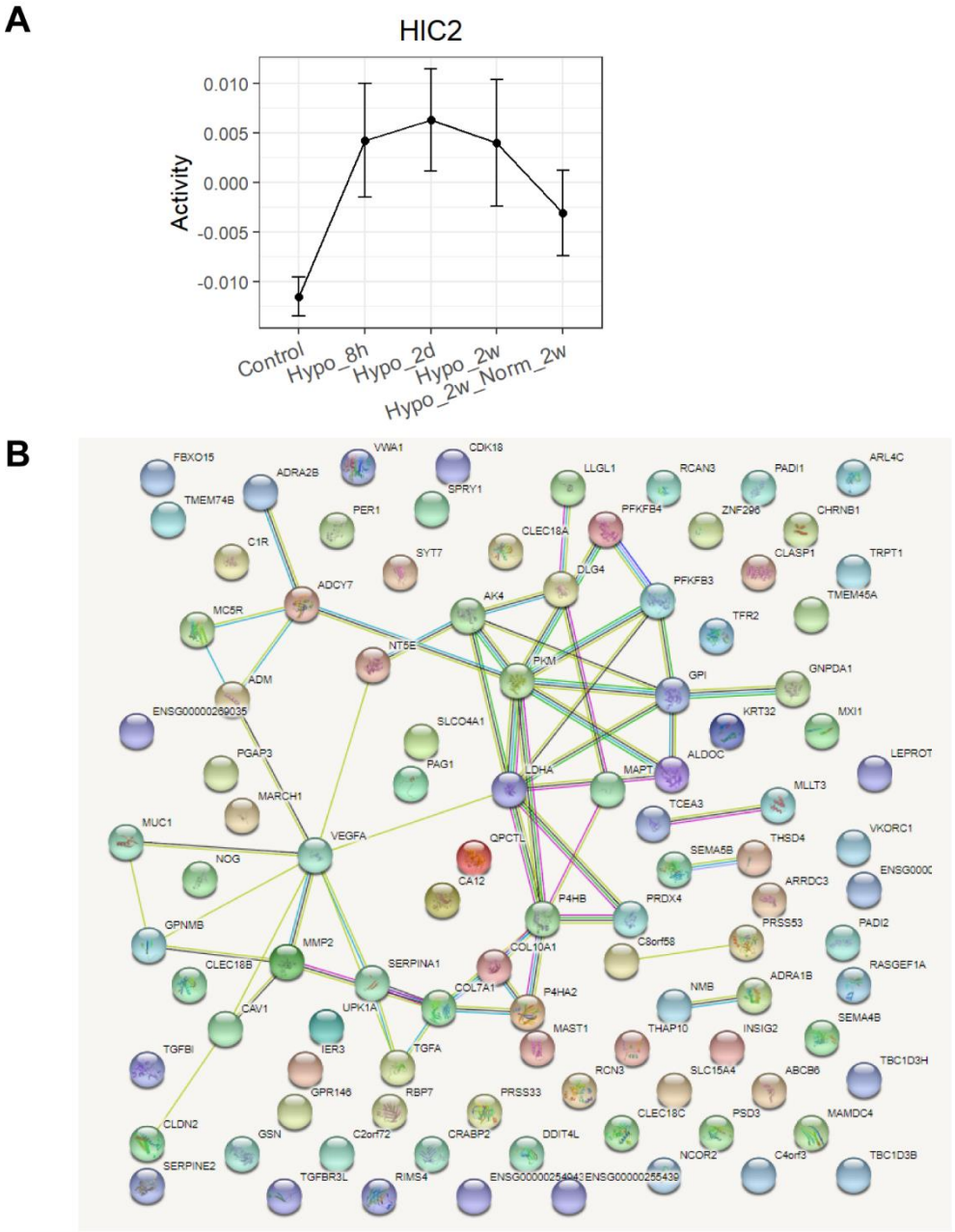

Figure 4.34 *ISMARA* found predicted HIC2 targets are regulated during hypoxia
A. Activity profile (mean + standard deviation) of HIC2 target genes. Hypo_8h is 8 hours hypoxia, Hypo_2d is 2 days hypoxia, Hypo_2w is 2 weeks hypoxia and Hypo_2w_Norm_2w is hypoxia recovery conditions. **B.** *STRING* representation of the predicted HIC2 targets that *ISMARA* found to be regulated in this dataset. Proteins with related interactions are connected by edges.

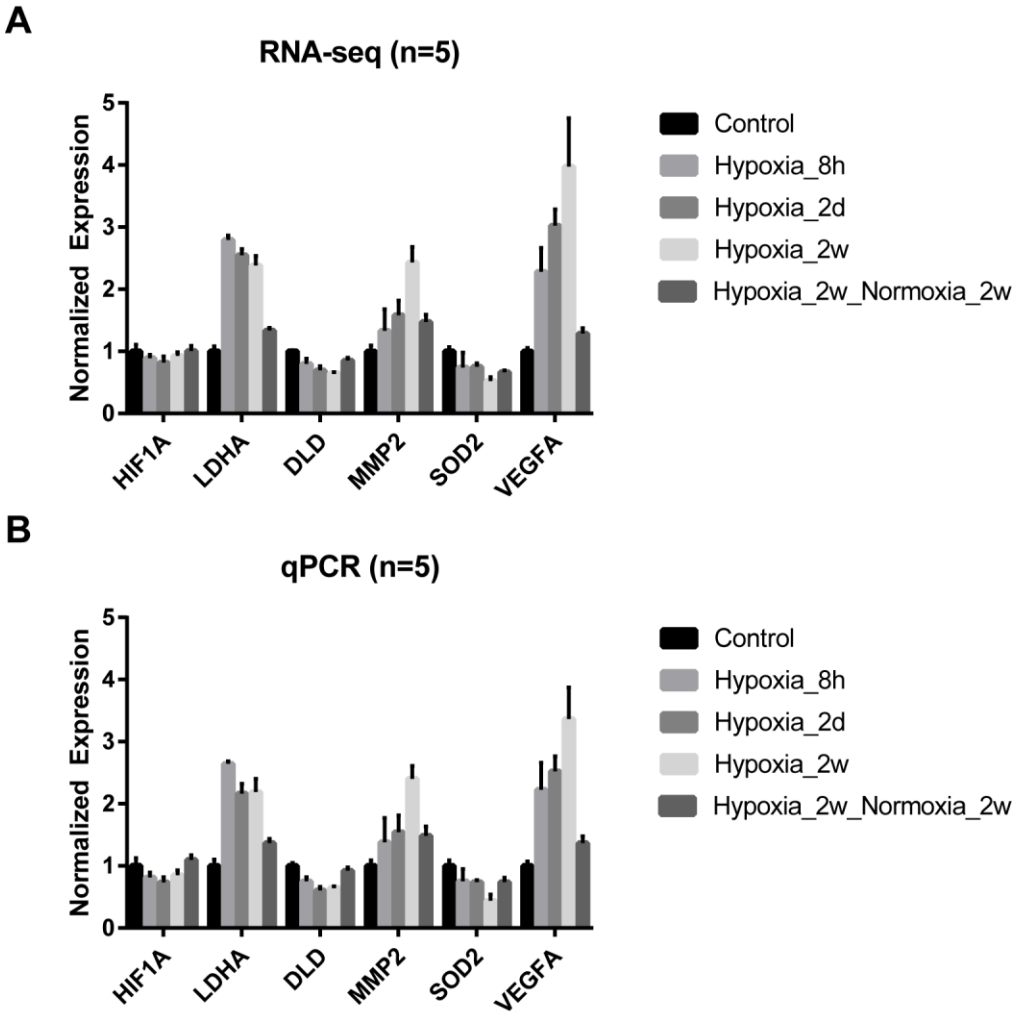

Figure 4.35 qPCR validation of selected RNA-seq results
Six genes shown by RNA-seq to be regulated during hypoxia were selected for qPCR validation. Expression values were normalized in both plots so that the average expression in control condition is 1. The mean and standard deviation are shown. Hypoxia_8h is 8 hours hypoxia, hypoxia_2d is 2 days hypoxia, hypoxia_2w is 2 weeks hypoxia, and hypoxia_2w_normoxia_2w is hypoxia recovery conditions. **A.** RNA-seq expression based on TPM values. **B.** qPCR expression based on ΔΔCt-derived fold changes.

Gene	Direction	5' to 3' Sequence
HIF1A	Forward	GAACGTCGAAAAGAAAAGTCTCG
HIF1A	Reverse	CCTTATCAAGATGCGAACTCACA
LDHA	Forward	ATGGCAACTCTAAAGGATCAGC
LDHA	Reverse	CCAACCCCAACAACTGTAATCT
DLD	Forward	CTCATGGCCTACAGGGACTTT
DLD	Reverse	GCATGTTCCACCAAGTGTTTCAT
MMP2	Forward	TACAGGATCATTGGCTACACACC
MMP2	Reverse	GGTCACATCGCTCCAGACT
SOD2	Forward	TTTCAATAAGGAACGGGGACAC
SOD2	Reverse	GTGCTCCCACACATCAATCC
VEGFA	Forward	AGGGCAGAATCATCACGAAGT
VEGFA	Reverse	AGGGTCTCGATTGGATGGCA
TUBB	Forward	TGGACTCTGTTCGCTCAGGT
TUBB	Reverse	TGCCTCCTTCCGTACCACAT

Table 4.1 qPCR primers

Trans. Factor	Z-score	Activity. diff Hypoxia_8h-Control	Activity. diff Hypoxia_2d-Control	Activity. diff Hypoxia_2w-Control	Activity. diff Hypoxia_2w_normoxia_2w-Control
MXI1_MYC_MYCN	5.985751	-0.04156	-0.02475	-0.01506	-0.00251
NRF1	4.709144	-0.01371	-0.01576	-0.01061	0.003826
HIF1A	4.702637	0.037304	0.031275	0.02326	0.004819
ARNT	3.969943	0.026974	0.026438	0.026203	0.00866
YBX1_FOS_NFYC_NFYA_NFYB_CEBPZ	3.583263	-0.0186	-0.01836	-0.01918	-0.00067
SIX5_SMARCC2_HCFC1	3.433916	-0.00965	-0.01465	-0.01192	-0.00198
PATZ1_KLF4	3.377575	0.005596	0.009211	0.009242	0.002927
EPAS1_BCL3	3.002331	0.026016	0.007556	0.006703	0.002857
ELF2_GABPA_ELF5	2.954512	-0.01664	-0.01427	-0.00911	-0.003
ZNF711_TFAP2A_TFAP2D	2.868172	-0.01185	-0.00687	-0.00262	-0.00193
PITX3	2.862096	-0.01573	-0.01699	-0.01735	-0.00546
HIC2	2.751785	0.015762	0.017797	0.015536	0.008411
WT1_MTF1_ZBTB7B	2.688741	0.013391	0.009849	0.008728	0.002619
AHR_ARNT2	2.687516	0.01195	0.011722	0.009293	-0.00062
BACH1_NFE2_NFE2L2	2.682164	-0.01493	-0.02065	-0.022	-0.00786
ESRRB_ESRRG	2.587262	-0.01033	-0.01781	-0.01954	-0.00744
CREB1	2.584654	-0.01806	-0.0142	-0.00827	0.000501
TAF1	2.576257	-0.02713	-0.02198	-0.01851	-0.00941
MECP2	2.561408	-0.00492	-0.00517	-0.00465	0.000575
SRF	2.516779	0.00486	-0.00354	0.02003	0.012833
MAZ_ZNF281_GTF2F1	2.50939	0.009488	0.011252	0.008038	0.000994
WRNIP1	2.468187	0.008925	0.005619	0.004495	0.000455
GTF2I	2.441592	0.012833	0.008224	0.007822	-0.00223
FOSL2_SMARCC1	2.405661	0.015183	0.016825	0.021105	0.007225
CEBPB	2.387667	0.000121	-8.50E-05	0.017481	-0.00211
ZFX	2.385102	-0.01729	-0.01044	-0.00985	-0.00739
FOSL1	2.366784	0.014858	0.014785	0.012922	0.001207
ERG	2.30491	0.013408	0.009604	0.005387	-0.00036

ATF4	2.302819	-0.0022	0.001443	0.018105	0.000298
ATF2_ATF1_ATF3	2.281693	-0.0125	-0.01138	-0.00888	-0.00064
HOXB2_UNCX_HOXD3	2.233088	-0.00791	-0.01757	-0.0108	-0.00452
ZNF143	2.229308	-0.01519	-0.0033	-0.00225	0.0039
MEIS2	2.147955	-0.00316	-0.00834	-0.01007	-0.00115
EBF1	2.10306	0.012099	0.012974	0.012471	0.000648
ZBTB14	2.088882	0.011221	0.004773	0.004971	-0.0014
PPARA	1.999801	-0.01269	-0.0133	-0.01742	-0.00929
HSF4	1.978446	-0.01271	-0.01112	-0.01423	-0.00162
NR5A2	1.959715	-0.01511	-0.01196	-0.01396	-0.00408
MYF6	1.952703	0.015941	0.013566	0.015877	0.009493
SOX4	1.946565	0.01024	0.011339	0.014389	0.000526
ETV1_ERF_FEV_ELF1	1.93076	-0.00905	-0.01416	-0.00812	0.005953
CTCF_CTCFL	1.920265	0.011528	0.011027	0.01426	0.004409
POU5F1_POU2F3	1.909801	0.012421	0.006181	-0.00168	0.001858
TFAP2C	1.895419	0.00527	0.006948	0.008185	-0.00154
MESP1	1.85942	0.009111	0.01138	0.010689	0.00149
MIXL1_GSX1_BSX_MEOX2_LHX4	1.811681	-0.00953	-0.01093	-0.01296	-0.00372
NFIA	1.811268	0.016127	0.005478	-0.00114	-0.00094
CEBPE_CEBPD	1.794327	0.009854	0.016299	0.00577	0.001329
NR1I2	1.786169	0.012905	0.010234	0.004668	0.000229
RFX7_RFX4_RFX1	1.7746	0.012971	0.000765	0.004503	2.90E-05

Table 4.2 Top 50 TF groups enriched

TFs are ordered by Z-score. Last four columns contain average target gene activity difference between control and each of Hypoxia_8h (8 hours hypoxia), Hypoxia_2d (2 days hypoxia), Hypoxia_2w (2 weeks hypoxia), and Hypoxia_2w_normoxia_2w (hypoxia recovery) conditions.

4.7 References

1. Structure and Function of the Eyes. Available at: https://www.merckmanuals.com/home/eye-disorders/biology-of-the-eyes/structure-and-function-of-the-eyes. (Accessed: 6th June 2015)
2. Antonetti, D. A. *et al.* Diabetic Retinopathy. *Diabetes* **55**, 2401–2411 (2006).
3. Strauss, O. The Retinal Pigment Epithelium in Visual Function. *Physiol. Rev.* **85**, 845–881 (2005).
4. An, E. *et al.* Secreted Proteome Profiling in Human RPE Cell Cultures Derived from Donors with Age Related Macular Degeneration and Age Matched Healthy Donors. *J. Proteome Res.* **5**, 2599–2610 (2006).
5. van Lookeren Campagne, M., LeCouter, J., Yaspan, B. L. & Ye, W. Mechanisms of age-related macular degeneration and therapeutic opportunities. *J. Pathol.* **232**, 151–164 (2014).
6. Otani, A. *et al.* Expressions of Angiopoietins and Tie2 in Human Choroidal Neovascular Membranes. *Invest. Ophthalmol. Vis. Sci.* **40**, 1912–1920 (1999).
7. dell'Omo, R. *et al.* Vitreous Mediators in Retinal Hypoxic Diseases. *Mediators Inflamm.* **2013**, (2013).
8. Matteucci, A. *et al.* Primary Retinal Cultures as a Tool for Modeling Diabetic Retinopathy: An Overview. *BioMed Res. Int.* **2015**, e364924 (2015).
9. Sakurai, E., Anand, A., Ambati, B. K., Rooijen, N. van & Ambati, J. Macrophage Depletion Inhibits Experimental Choroidal Neovascularization. *Invest. Ophthalmol. Vis. Sci.* **44**, 3578–3585 (2003).
10. Simó, R., Sundstrom, J. M. & Antonetti, D. A. Ocular Anti-VEGF Therapy for Diabetic Retinopathy: The Role of VEGF in the Pathogenesis of Diabetic Retinopathy. *Diabetes Care* **37**, 893–899 (2014).
11. Li, S.-Y., Fu, Z. J. & Lo, A. C. Y. Hypoxia-Induced Oxidative Stress in Ischemic Retinopathy. *Oxid. Med. Cell. Longev.* **2012**, e426769 (2012).
12. Stefánsson, E., Geirsdóttir, A. & Sigurdsson, H. Metabolic physiology in age related macular degeneration. *Prog. Retin. Eye Res.* **30**, 72–80 (2011).
13. Blasiak, J. *et al.* Oxidative Stress, Hypoxia, and Autophagy in the Neovascular Processes of Age-Related Macular Degeneration. *BioMed Res. Int.* **2014**, e768026 (2014).
14. Barben, M. *et al.* Hif1a inactivation rescues photoreceptor degeneration induced by a chronic hypoxia-like stress. *Cell Death Differ.* 1 (2018). doi:10.1038/s41418-018-0094-7
15. Kurihara, T. *et al.* Hypoxia-induced metabolic stress in retinal pigment epithelial cells is sufficient to induce photoreceptor degeneration. *eLife* **5**,
16. Forooghian, F., Razavi, R. & Timms, L. Hypoxia-inducible factor expression in human RPE cells. *Br. J. Ophthalmol.* **91**, 1406–1410 (2007).
17. Dunn, K. C., Aotaki-Keen, A. E., Putkey, F. R. & Hjelmeland, L. M. ARPE-19, a human retinal pigment epithelial cell line with differentiated properties. *Exp. Eye Res.* **62**, 155–169 (1996).

18. Mousa, S. A., Lorelli, W. & Campochiaro, P. A. Role of hypoxia and extracellular matrix-integrin binding in the modulation of angiogenic growth factors secretion by retinal pigmented epithelial cells. *J. Cell. Biochem.* **74**, 135–143 (1999).
19. Jiang, Z. *et al.* Ribosome profiling reveals translational regulation of mammalian cells in response to hypoxic stress. *BMC Genomics* **18**, 638 (2017).
20. Samuel, W. *et al.* Appropriately differentiated ARPE-19 cells regain phenotype and gene expression profiles similar to those of native RPE cells. *Mol. Vis.* **23**, 60–89 (2017).
21. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
22. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
23. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
24. Jiang, H. & Zhan, T. Unit-Free and Robust Detection of Differential Expression from RNA-Seq Data. *Stat. Biosci.* **9**, 178–199 (2017).
25. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
26. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
27. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
28. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
29. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
30. Tenenbaum, D. KEGGREST: Client-side REST access to KEGG. *R package version 1.20.0* Available at: http://bioconductor.org/packages/KEGGREST/. (Accessed: 26th September 2018)
31. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).
32. Shao, Z. ERSSA: Empirical RNA-seq Sample Size Analysis. *Bioconductor* Available at: http://bioconductor.org/packages/ERSSA.
33. Balwierz, P. J. *et al.* ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.* (2014). doi:10.1101/gr.169508.113
34. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447-452 (2015).
35. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
36. CaltechDATA. *Caltech Library* Available at: https://www.library.caltech.edu/caltechdata. (Accessed: 9th December 2018)

37. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* **25**, 402–408 (2001).
38. Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* **40**, D1144–D1149 (2012).
39. Tafani, M. *et al.* The Interplay of Reactive Oxygen Species, Hypoxia, Inflammation, and Sirtuins in Cancer Initiation and Progression. *Oxidative Medicine and Cellular Longevity* (2016). doi:10.1155/2016/3907147
40. Krock, B. L., Skuli, N. & Simon, M. C. Hypoxia-Induced Angiogenesis Good and Evil. *Genes Cancer* **2**, 1117–1133 (2011).
41. GeneCards - Human Genes | Gene Database | Gene Search. Available at: https://www.genecards.org/. (Accessed: 16th October 2018)
42. Sharrocks, A. D. The ETS-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.* **2**, 827–837 (2001).
43. Aprelikova, O., Wood, M., Tackett, S., Chandramouli, G. V. R. & Barrett, J. C. Role of ETS transcription factors in the hypoxia-inducible factor-2 target gene selection. *Cancer Res.* **66**, 5641–5647 (2006).
44. Qiao, N. *et al.* Ets-1 as an early response gene against hypoxia-induced apoptosis in pancreatic *β*-cells. *Cell Death Dis.* **6**, e1650 (2015).
45. Dykes, I. M., van Bueren, K. L. & Scambler, P. J. HIC2 regulates isoform switching during maturation of the cardiovascular system. *J. Mol. Cell. Cardiol.* **114**, 29–37 (2018).
46. Lv, L. *et al.* The DNA methylation-regulated miR-193a-3p dictates the multi-chemoresistance of bladder cancer via repression of SRSF2/PLAU/HIC2 expression. *Cell Death Dis.* **5**, e1402 (2014).
47. Dykes, I. M. *et al.* HIC2 is a novel dosage-dependent regulator of cardiac development located within the distal 22q11 deletion syndrome region. *Circ. Res.* **115**, 23–31 (2014).
48. Costa, C., Incio, J. & Soares, R. Angiogenesis and chronic inflammation: cause or consequence? *Angiogenesis* **10**, 149–166 (2007).
49. Tas, S. W., Maracle, C. X., Balogh, E. & Szekanecz, Z. Targeting of proangiogenic signalling pathways in chronic inflammation. *Nat. Rev. Rheumatol.* **12**, 111–122 (2016).
50. Bodamyali, T., Stevens, C. R., Billingham, M. E. J., Ohta, S. & Blake, D. R. Influence of hypoxia in inflammatory synovitis. *Ann. Rheum. Dis.* **57**, 703–710 (1998).
51. Lim, L. S., Mitchell, P., Seddon, J. M., Holz, F. G. & Wong, T. Y. Age-related macular degeneration. *The Lancet* **379**, 1728–1738 (2012).
52. Stitt, A. W., Lois, N., Medina, R. J., Adamson, P. & Curtis, T. M. Advances in our understanding of diabetic retinopathy. *Clin. Sci.* **125**, 1–17 (2013).
53. Liu, L. & Simon, M. C. Regulation of transcription and translation by hypoxia. *Cancer Biol. Ther.* **3**, 492–497 (2004).
54. van den Beucken, T. *et al.* Translational control is a major contributor to hypoxia induced gene expression. *Radiother. Oncol.* **99**, 379–384 (2011).

55. Axon Guidance Signaling Modulates Epithelial to Mesenchymal Transition in Stem Cell-Derived Retinal Pigment Epithelium | bioRxiv. Available at: https://www.biorxiv.org/content/early/2018/02/13/264705. (Accessed: 1st November 2018)
56. Hirasawa, M. *et al.* Transcriptional factors associated with epithelial-mesenchymal transition in choroidal neovascularization. *Mol. Vis.* **17**, 1222–1230 (2011).

Shao Z and Kornfield J. "ERSSA: Empirical RNA-seq Sample Size Analysis." *R package version 0.99.7* (2018). DOI: 10.18129/B9.bioc.ERSSA.

Chapter 5. ERSSA, a subsampling-based RNA-seq sample size analysis method

5.1 Introduction

For comparative RNA-seq experiments, selecting the appropriate sample size is an important optimization step[1]. Comparisons with limited biological replicates tend to be under-powered with inferior differential expression (DE) discovery. On the other hand, the improvement in DE identification becomes marginally small as more replicates are added, while cost increases linearly[2]. In the past few years, several algorithms have been proposed to identify the appropriate sample size for RNA-seq experiments[3]. As summarized in **Chapter 1.4**, a recent survey article set out to systematically evaluate these algorithms and found all of them are ineffective in estimating the correct sample size when the inter-condition differences (effect size) are modest[4]. Here, we present a new method called Empirical RNA-seq Sample Size Analysis (ERSSA[5]) that tackles this challenge from a different angle: instead of attempting to predict the appropriate sample size using *a priori* assumptions prior to acquiring data, ERSSA evaluates existing RNA-seq data to estimate the marginal return in DE gene discovery as sample size increases. The results then allow the user to assess whether the number of replicates already acquired is sufficient for their study.

To demonstrate ERSSA's utility, we applied the analytical approach to several publicly available RNA-seq datasets that span four types of experimental systems: mouse tissue,

human tissue, human population, and *in vitro* cell culture. We will cover comparisons where ERSSA shows sufficient replicates have been included and examples where additional samples could be beneficial. More broadly, the results from this study depict the importance of post-data acquisition sample size validation, a practice largely ignored in the RNA-seq research community. By incorporating ERSSA (and the accompany software, *ERSSA*) into existing research projects, users can benefit from both a better understanding of their experimental dataset and avoid biological interpretation based on insufficiently powered experiments.

5.2 Materials and Methods

Software development and data acquisition

The ERSSA algorithm is implemented in the statistical programming language R and the software package (also named *ERSSA*) is available on Bioconductor for download[5]. *ERSSA* currently supports differential expression testing with *edgeR*[6] and *DESeq2*[7], but additional tools can be incorporated through manual modification of the code[8].

To demonstrate *ERSSA's* utility, we applied the technique to four biomedically-relevant datasets that represent a diverse set of RNA-seq experiments:

1) GTEx: a dataset containing human heart and muscle tissues[9].

2) Bottomly: brain tissues from two mouse strains often used in neuroscience research[10].

3) Montgomery-Pickrell (MP): lymphoblastoid cell lines derived from Europeans and Nigerians[11,12].

4) Fossum: cell lines treated with transcription factor knockdown or control siRNA[13].

The four publicly available datasets were downloaded from the recount and recount2 projects, which are systematic efforts to generated gene expression count table from thousands of RNA-seq studies[14,15]. Bottomly and MP datasets were obtained from the original recount project. GTEx and Fossum datasets were downloaded from the recount2 project. Recount2 count files were scaled using the *recount* software package as recommended by the authors before usage in *ERSSA*[16]. All analyses were performed in R with random seed set to 1. For GTEx and MP, the entire datasets are too large for the purpose of *ERSSA* evaluation. For these two datasets, 25 replicates of each condition were randomly selected for analysis. Furthermore, the GTEx dataset is broken down into three datasets containing 5, 10, and 25 replicates per condition for testing purposes. **Table 5.1** (GTEX)**, 5.2** (GTEX)**, 5.4** (Bottomly)**, 5.6** (MP)**, and 5.8** (Fossum) display the original ID of the samples used in this study.

Algorithm

Before running *ERSSA*, two inputs need to be prepared ahead of time:

1) A RNA-seq count table containing samples from two conditions;

2) A table that specifies the sample-condition relationship.

Examples of the input tables are included in the *ERSSA* vignette section 3.2[8]. Once the inputs are ready, the ERSSA algorithm runs through four sequential steps: 1) filter count table to remove non-expressed genes, 2) generate unique subsamples (combinations of samples from both conditions) at each replicate level, 3) perform DE analysis of the subsamples using well-established DE software, and 4) generate plots to visualize and interpret the results. Here, we will briefly describe each step; full descriptions are available in the *ERSSA* software documentation[5].

Filter count table

Removal of non-expressed genes before differential expression comparison is a widely accepted practice that is recommended to maximize discovery[17]. *ERSSA* uses a simple gene-wise average Count per Million (CPM) value to identify non-expressed genes:

$$Average\ CPM_g = \left(\sum_i^M \frac{count_{gi} * 10^6}{total\ count_i}\right) / M\ ,$$

where g denotes a specific gene, i denotes a specific sample, and M is the total number of samples. At default, genes with average CPM<1 are removed from further analysis. Additionally, the user can elect to supply their own pre-filtered count table generated from other filtering methods. The filtered count table is then passed to the differential testing step (*DE analysis*, below).

Generate subsample combinations

One of the core features of ERSSA is to use subsamples selected at each replicate level to observe the trend in DE gene discovery as a function of sample size. Using the sample-condition relationship table, the algorithm first identifies the smaller of the two numbers of replicates (*N*) available among the two conditions. Next the algorithm performs a pre-determined number of subsampling, where possible, at each step-wise replicate levels (*n*= 2, 3, 4, ..., *N*-1). For large *N*, the number of unique subsamples at some replicate levels *n* becomes very large and it is not necessary to evaluate all combinations. Typically, N_s=30-50 subsamples at each replicate level are sufficient to observe the pattern in DE gene discovery as a function of sample size. When the total number of unique sample combinations is less than N_s (for example: *N*=3 and *n*=2 produces 9 unique subsamples), then all unique subsamples are selected. When the number of possible combinations exceeds N_s, the algorithm generates N_s unique and random subsamples. At default, *ERSSA* generates N_s=30 subsamples at each *n*; the user can change the value N_s, which affects run time linearly. The generated subsample combinations together with the filtered count table are then passed to DE software for statistical testing.

DE analysis

Using the previously generated filtered count table and subsample combinations, *ERSSA* makes a call to a DE software for statistical testing. Currently, *ERSSA* supports *edgeR* and *DESeq2*, two of the most widely used DE software packages, for differential discovery[18]. At default, *edgeR* is used as it is slightly faster in runtime compare to *DESeq2*. After statistical

testing, DE genes are defined by two cutoff statistics: multiple-test corrected p-value (adjusted p-value in *DESeq2* and False Discovery Rate (FDR) in *edgeR*)<0.05 and |log2-fold-change|>1. Unless otherwise specified, genes that match both criteria are considered to be DE in the scope of ERSSA calculations. At the end of the test, the identity of the DE genes are retained and subsequently used for quantification and visualization.

Plot DE results

Using the test outputs, *ERSSA* generates four summary plots to help the user interpret the results. Further detail regarding the four summary plots and their utility can be found in the Results section (**Chapter 5.3**) below. The first plot displays the number of DE genes identified as a function of the number of replicates (*n*) employed. The second plot shows the percent change in average number of DE genes identified as *n* increases. The third plot displays the number of DE genes that are identified in all of the subsamples at each replicate level, revealing the increase in the consistency of DE gene discovery as *n* increases. Based on the statistics from this plot, the percentage of consistently discovered DE genes found in all subsamples can be calculated. The last plot must be viewed with care. Since the ground truth is not known in RNA-seq studies, the DE discovery at the full replicate level is used as the "pseudo ground truth" for calculating a True Positive Rate (TPR) and False Positive Rate (FPR) of DE discovery for each subsample, color coded for corresponding sample size *n*. The average rates (TPR and FPR) at each replicate level are also displayed. It is important to consider that when the discovery with the full dataset is not reliable, interpretation of this

plot is inappropriate. Only when *ERSSA* shows the full dataset is able to detect a majority of DE genes consistently (through interpretation of first three plots) is this analysis useful.

All four plots are created using the *ggplot2* R package with the plot objects and the plotted data saved[19]. From there, further modifications can be applied to adapt to various display preferences. Lastly, the raw data generated during the calculations are also available to the user for additional analysis outside the scope of *ERSSA*.

5.3 Results

GTEx – human muscle and heart tissue dataset

The original dataset from GTEx contains hundreds of human tissue samples, including ones from muscle and heart. For the purpose of *ERSSA* evaluation, 25 muscle and heart samples were randomly selected for inclusion (**Tables 5.1, 5.2**). To illustrate how ERSSA enables discrimination of statistical power as a function of the number of replicates, we analyzed three cases: N=5, N=10, and N=25. A single random selection of 5 replicates taken from the 25 total illustrates what would be observed if N=5. Likewise, a single random selection of 10 from the 25 total illustrates what would be observed with N=10. This example shows how a user can evaluate a small number of replicates and see the likelihood that significant gains could be achieved by increasing the number of replicates (results for N=5 indicate that additional replicates could bring significant improvements, which is confirmed by the

analysis for N=10). This example also shows how a user can evaluate an intermediate number of replicates and accurately infer that further replicates are unlikely to reveal additional DE genes (results for N=10 suggest that the improvement with increasing N has saturated, which is confirmed by the analysis for N=25).

For most studies, the average number of identified DE genes generally increases as sample size increases. In our literature search, we observed that most RNA-seq studies use $N \leq 5$ replicates. While this might be sufficient in well controlled studies testing large transcriptome changes, others could see significant improvement in detection with more replicates. Therefore, we chose a mock experiment with only 5 replicates as an example. In this analysis, the test result with all 5 replicates is taken as the "pseudo ground truth" and the improvement with increasing n is computed for subsamples of n=2, 3, and 4. The increase in the mean number of DE genes and the decrease in the variability among the different subsamples is strong over this range of n (**Figure 5.1A**, note the increase in the mean (red line) and the median, as well as the decrease in the difference from the 1st to 3rd quartiles as n increases from 2 to 4). The potential for significant further increase in the number of DE genes discovered is indicated by the percent increase in the DE gene discovery for n=5 relative to n=4 (**Figure 5.1B** suggests improvements in the order of 5-10% could be achieved with additional replicates). Furthermore, *ERSSA* shows that there is potentially a significant increase in the fraction of genes that can be reliably discovered with more replicates (**Figure 5.2A**) as only 40% of the DE genes at n=4 is found in all subsample results (**Table 5.3**).

Finally, the graph of TPR and FPR must be regarded with caution in the view of the previous three *ERSSA* outputs: the "pseudo ground truth" using *N*=5 is not a good approximation for the ground truth (**Figure 5.2B**).

Based on the *ERSSA* results for *N*=5, a user might choose to acquire 5 additional replicates to improve discovery. As anticipated based on **Figure 5.1**, the *N*=10 dataset identified around 6200 DE genes (**Figure 5.3A**), a significant increase relative to around 5500 for *N*=5. Indeed, 5 additional replicates provided a DE gene discovery increase in the low double digits percentage range, which is consistent with the prediction base on the *N*=5 analysis. With *N*=10, *ERSSA* also suggests that the number of DE genes identified hardly changed with increase of *n* from 8 to 9 and 10 (**Figure 5.3A and B**). This observable "plateau" at the large *n* regime suggests that a majority of DE genes have already been discovered with the available biological replicates. At this regime, the additional benefit of increasing replicates is the confidence that regardless of the particular set of samples obtained for the experiment, each one will find DE genes that would be identified in any other subsamples (**Figure 5.4A, Table 5.3**).

Based on the evidence that the "pseudo ground truth" represented by the DE genes discovered using *N*=10 is a good approximation, the fourth output of *ERSSA* has greater value in this case than it did for *N*=5. The graph of TPR and FPR indicates that the FPR only begins to decrease when $n > 4$ and that it decreases from around 5% at *n*=6 to around 2% for *n*=9 (**Figure 5.4B**). Additionally, the TPR continuously progress from *n*=8 to 9 to 10, suggesting

that the actual ground truth might reveal that TPR for *N*=10 is less than 100%. Lastly, it is again worth noting that this particular sampling of *N*=10 replicates is *not* a perfect representation of the ground truth; the data point for the full data set is placed at TPR=1, FPR=0 simply because it is used as the "pseudo ground truth."

When we proceed to *N*=25 replicates, the expectations based on application of *ERSSA* at *N*=10 are borne out. The mean number of genes discovered hardly changed for *n*>10 (**Figure 5.5A and B**). The benefit of the increase in the number of replicates is that approximately 5300 DE genes are reliably found by any subsamples at *n*=24 (**Figure 5.6A**), corresponding to a high of 83% of all DE genes discovered (**Table 5.3**). The TPR increases a few percent as *n* increases from 21 to 24 and the FPR is consistently less than 2% for *n*>21 (**Figure 5.6B**), suggesting that the marginal return on acquiring additional replicates has become increasingly small.

As demonstrated with the GTEx example dataset, *ERSSA* helps the user deal with the fact that while the true list of DE genes is typically never known, the existing replicates at hand can be used to predict the amount of improvement to be gained by acquiring additional replicates. For *N*=5, *ERSSA* reveals that the discovery is most likely incomplete (**Figure 5.1 and 5.2A**), in which the TPR and FPR calculated are not useful (**Figure 5.2B**). *ERSSA* analysis of *N*=5 replicates correctly anticipated that an additional 5 replicates could bring a substantial improvement (compare **Figure 5.1A and 5.2A**, to **Figure 5.3A and 5.4A**). We

also noticed that a large number of replicates is required to obtain a good approximation of the true list of DE genes (average TPR>95% for *N*=25, **Figure 5.6B**).

Bottomly – mouse tissue dataset

The Bottomly dataset contains brain samples from two mouse strains that are commonly used in neuroscience research (**Table 5.4**). When all 10 replicates available were tested with *ERSSA*, it appears that there are relatively few DE genes between the two strains (**Figure 5.7A**). Consistent with previous observation with the GTEx dataset, there is a clear pattern of improved DE gene discovery as additional replicates are used (**Figure 5.7B**). Some subsamples at the low *n* regime did found significantly more DE genes that pass the FDR cutoff. However, these results are outliers with many false positive discoveries as indicated by the relatively low number of DE gene discovery with *n=10*. After inclusion of six replicates, the improvement is primarily in the fidelity of the DE gene identification, rather than the number of DE genes identified (compare the increase in bars and mean DE gene curve for *n*>6 in **Figure 5.8A**). Based on the fraction of consistently identified DE genes (the consensus DE genes represent 42% of the putative DE genes identified in at least one of the subsamples at *n*=9, **Figure 5.8A, Table 5.5**), the discovery with the full dataset is likely an imperfect approximation for the true DE gene list. This must be kept in mind when viewing the fourth output of *ERSSA*, where the full dataset is used as the "pseudo ground truth". In this dataset, TPR calculation shows a strong improvement from *n*=6 to 9 in **Figure 5.8B,** suggesting that there is much to be gained by increasing the number of replicates.

MP – human population dataset

The MP dataset is collected as part of the International HapMap Project and contains lymphoblastoid cell lines derived from European and Nigerian individuals. The entire dataset contains more than 60 replicates per group. To illustrate the use of *ERSSA*, 25 replicates from each population were randomly selected for testing (**Table 5.6**). Compared to the GTEx and Bottomly datasets, *ERSSA* shows that DE gene discovery in the MP dataset requires a much greater number of replicates (compare **Figure 5.9** to **Figures 5.5 and 5.7**). As *n* increases, the mean number of putative DE genes gradually increases and reaches plateau at around *n*=18 with approximately 1000 putative DE genes on average. However, fewer than 400 of those identified at *n*=18 are consistently found in all subsamples that were analyzed (**Figure 5.10A**). Indeed, consensus among the subsamples remains quite low across all sample sizes tested (even at *n*=24, there are only 720 consensus DE genes, representing 53% of the putative DE genes identified in at least one of the subsamples, **Table 5.7**). Similarly, the average TPR is quite low in the low replicate regime and gradually improves to at least 0.75 as *n* increases to 14 and higher (**Figure 5.10B**). However, even at the high replicate regime (*n*=20 to 24), the TPR continues to increase significantly with additional replicates. Thus, *ERSSA* shows that while the total number of DE genes detectable has plateaued, a more accurate DE discovery could be obtained by increasing the number of replicates.

Fossum – cell line dataset

The Fossum dataset originated from an effort to identify the transcriptome change after EHF transcription factor perturbation. In this dataset, EHF expression was knocked down using siRNA and a separate cohort of samples were treated with negative control siRNA. In both conditions, five cell culture replicates are available for *ERSSA* (**Table 5.8**). This example illustrates an experiment for which the default setting of the fold-change threshold (|log2-fold-change|>1) is too stringent. Using the default settings, only a few DE genes are detected at all available replicate levels (**Figure 5.11**), which suggests that treatment-induced transcriptome changes are mild and might be better explored by relaxing the fold change cutoff to 0.5. Analysis with the relaxed parameter allows a larger number of putative DE genes to be identified and the results show a substantial increase in the number of DE gene discovery as sample size increases, suggesting the plateau has not been reached (**Figure 5.12**). Robustness among the subsamples also remains low with consensus DE genes at n=4 representing only 18% of putative DE genes found in at least one of the subsamples (**Figure 5.13A, Table 5.9**). Based on these observations, the fourth plot of *ERSSA* is of limited use as the "pseudo ground truth" with N=5 is likely a poor representation of the real ground truth. Indeed, the primary message from **Figure 5.13B** is that increasing the number of replicates brings a dramatic increase in calculated TPR, providing strong motivation to acquire additional replicates.

5.4 Discussion

In recent years, many algorithms have been proposed to guide the selection of the appropriate sample size for RNA-seq experiments[4]. These algorithms typically use pilot or publicly available RNA-seq data to estimate parameters and rely on *a priori* assumptions about the effect size to estimate the sample size. Some well cited packages that use this approach include *Scotty* (2013)[20], *RNASeqPowerCalculator* (2014)[21], *ssizeRNA* (2016)[22], and *powsimR* (2017)[23]. Typically, these algorithms are applied prior to RNA-seq experiments to estimate the sample size needed to gain sufficient statistical power. However, a recent paper showed that these approaches actually perform poorly when the effect size is modest[4], as is the case for a large fraction of RNA-seq experiments. To the best of our knowledge, there are no methods available that are specifically designed to validate sample size post-RNA-seq sample acquisition.

The ERSSA algorithm and the accompanying *ERSSA* package address these needs. In ERSSA, a subsampling approach is used with existing DE analysis software to determine whether the appropriate sample size has been reached. Here, we applied the analytical approach to several publicly available datasets and found favorable utility across different experimental settings. In three datasets (GTEx, Bottomly and MP), *ERSSA* shows sufficient biological replicates have been included for DE gene discovery (**Figure 5.5, 5.7, and 5.9**). Furthermore, we performed the GTEx analysis at three maximum sample sizes and verified that *ERSSA* forecast correctly from N=5 that a substantial increase in the number of putative

DE genes could be achieved by obtaining additional replicates and from *N*=10 that a significant increase in the number of consensus DE genes could be obtained with additional replicates (**Figure 5.1-5.6**). In one study (Fossum), the results suggest additional samples could improve the number of DE gene discovery (**Figures 5.11 to 5.13**). Although obtaining additional replicates may not always be feasible or economical, ERSSA enables the user to appreciate the limitations of their dataset.

In addition to estimating the quantity of putative DE genes at a certain replicate size, ERSSA's subsampling approach also enables analysis of the fidelity of DE gene discovery. Using the number of consensus DE genes across all subsamples as a metric, we found that the four studies typically have low consistency at the low replicate regime and a continuous improvement as sample size increases (**Figure 5.6A, 5.8A, 5.10A, 5.13A**). The lack of consistency with small *n* is likely due to a combinations of limited DE gene discovery and higher variability. On the other hand, it is interesting to note that even at the high replicate regime, we did not observe a plateau in the consensus/putative ratio of DE genes (**Table 5.3, 5.5, 5.7, 5.9**). Instead, we consistently observed a continuous increase in the percentage of consensus DE genes calculated from the subsample analyses.

TPR and FPR are other metrics of evaluating the variability in the DE gene discovery. In figures plotting the two rates, we found that the four studies behaved very differently in terms of their TPR/FPR profile. In three studies (Bottomly, MP and Fossum), the average FPRs are across the board quite low, typically in the range of 0.01 to 0.04 (**Figure 5.8B, 5.10B, 5.13B**).

In the GTEx study, average FPRs are consistently higher in all three analyses, but still reasonable within a 0.05 to 0.10 range (**Figure 5.2B, 5.4B, 5.6B**). Additionally, we found that while TPR tend to increase consistently as a function of sample size, FPR can initially increase and then decrease (**Figure 5.4B, 5.6B, 5.10B**). This interesting FPR pattern is clearly observable in both GTEx and MP datasets, suggesting that while initial sample size addition improves discovery, it can also increase the number of false positives. Another interesting observation is that while TPR improvement appears monotonic, its marginal increment becomes smaller as sample size becomes larger. This is especially noticeable in the three larger datasets with $N>5$, particularly in the $N=25$ GTEx dataset (**Figure 5.6B**). Lastly, we would like to again emphasize that the rates are calculated by setting the list of DE genes discovered with the full available dataset as the ground truth. In the datasets where discovery is under-powered, the calculated rates are unreliable and no conclusions should be drawn.

While the ERSSA algorithm does not use any *a priori* assumptions about the dataset, there are several runtime specifications that need to be considered. First, the number of subsamples to run per replicate level is an important *ERSSA* parameter. For the majority of datasets, we found 30 subsamples are sufficient to provide a good estimate of the discovery. On the other hand, some datasets are prone to outliers, where additional subsamples could reduce the sensitivity to any particular result (*e.g.* 50 subsamples used in MP dataset). Two additional runtime specifications that can be adjusted are the statistical test and fold change cutoffs used

for DE consideration. In particular, depending on the effect side, the fold change cutoff can be either tightened or relaxed to fit the comparison. We saw an example of this in the Fossum dataset where the transcriptome change after treatment is quite mild, so the fold change cutoff is relaxed to increase DE gene discovery (**Figure 5.11, 5.12**).

While using *ERSSA*, we also found it useful to run the algorithm multiple times with different random seeds. Since *ERSSA* uses a random process to generate the subsamples, running the algorithm multiple times causes it to generate different groups of subsamples. When the conclusions are consistent across multiple simulations, the user can gain confidence that the results are independent of the specific choice of subsamples. Lastly, depending on the size of the dataset, hundreds to thousands of DE comparisons may need to be run (*e.g. N*=25 with default 30 subsamples require a total of 691 statistical tests). To reduce run time, the user is highly encouraged to turn on parallel processing in *ERSSA*, which is managed by the *BiocParallel* package[24].

5.5 Conclusions

For any comparative RNA-seq experiment, selecting the appropriate sample size is an important and challenging design decision. ERSSA is a novel algorithm that enables users to determine whether an existing RNA-seq dataset has sufficient number of biological replicates. In this study, we applied an implementation of the algorithm in R to four publicly available RNA-seq datasets to demonstrate the utility of this approach. With these studies,

ERSSA shows good performance with a diverse set of experimental conditions including different sample size, sample type, and effect size. Results from our study further demonstrate that post-RNA-seq sample size analysis is an important quality control step. The *ERSSA* software can be broadly applied to any simple comparative RNA-seq studies (meaningfully with $N>3$) and is available for download at the Bioconductor website.

5.6 Figures and Tables

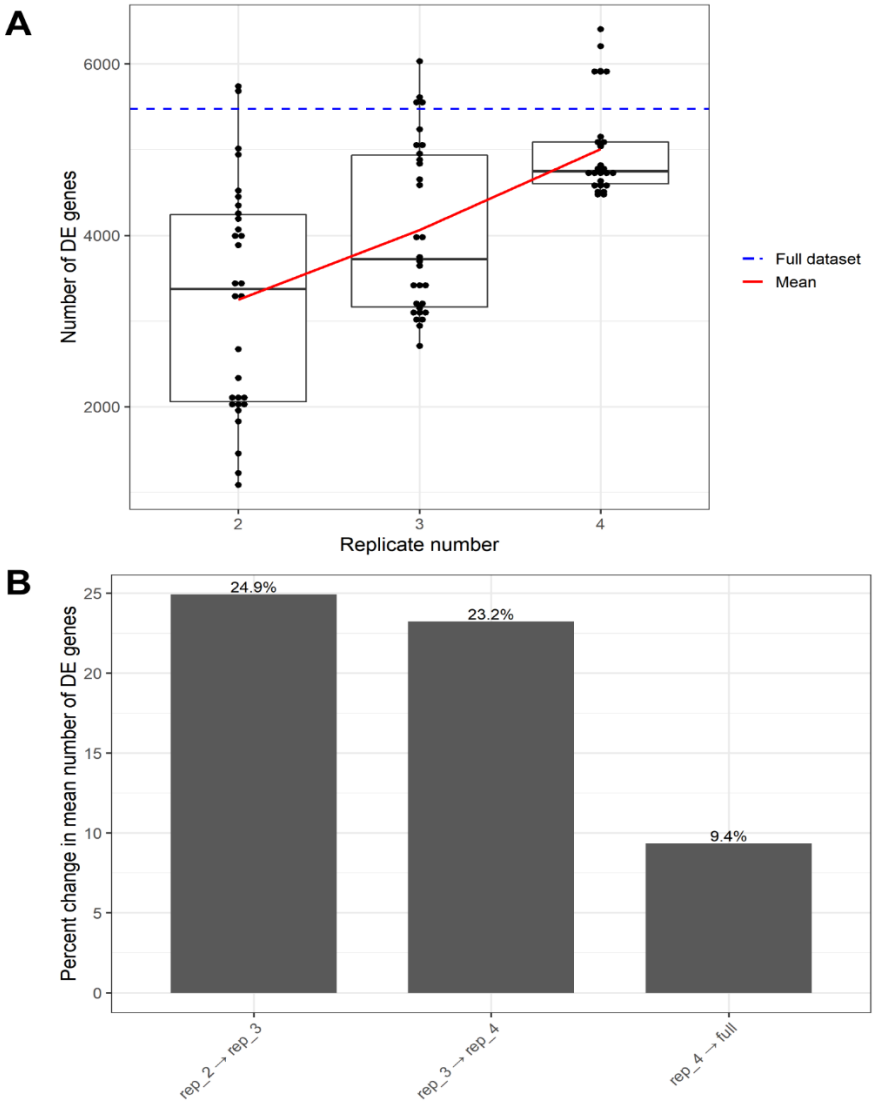

Figure 5.1 *ERSSA* with 5 replicate GTEx dataset – part 1
Plots generated by running *ERSSA* with 5 replicate GTEx dataset comparing human muscle and heart tissues. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Dots represent the number of DE genes discovered in each subsample. Boxplot summarizes the results. Red and blue lines represent the average and full discovery, respectively. **B.** Percent change in average DE gene discovery as sample size increases.

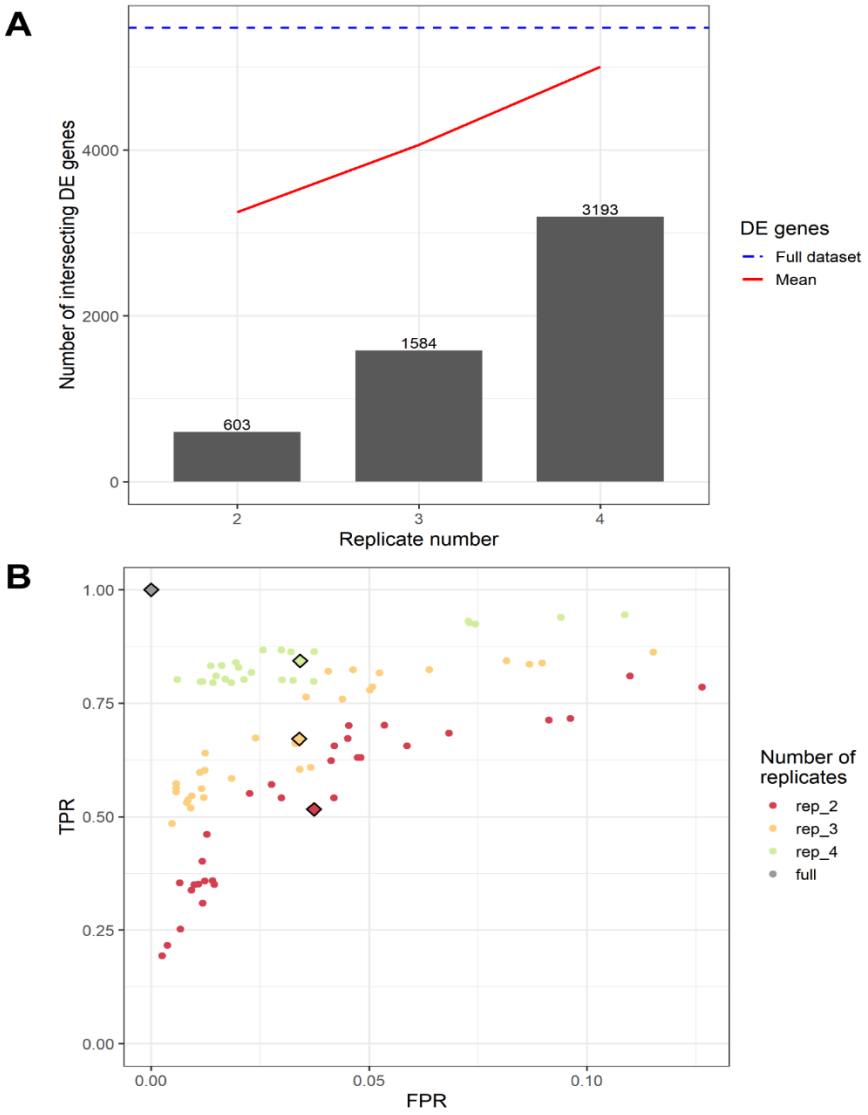

Figure 5.2 *ERSSA* with 5 replicate GTEx dataset – part 2
Plots generated by running *ERSSA* with 5 replicate GTEx dataset comparing human muscle and heart tissues. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Bar plot of the number of DE genes found in all subsamples at each replicate level. Red and blue lines represent the average and full discovery, respectively. **B.** Dots indicate FPR and TPR of each subsample. Diamonds represent the average rates at each replicate level. DE discovery with the full dataset is set as the ground truth for the calculations.

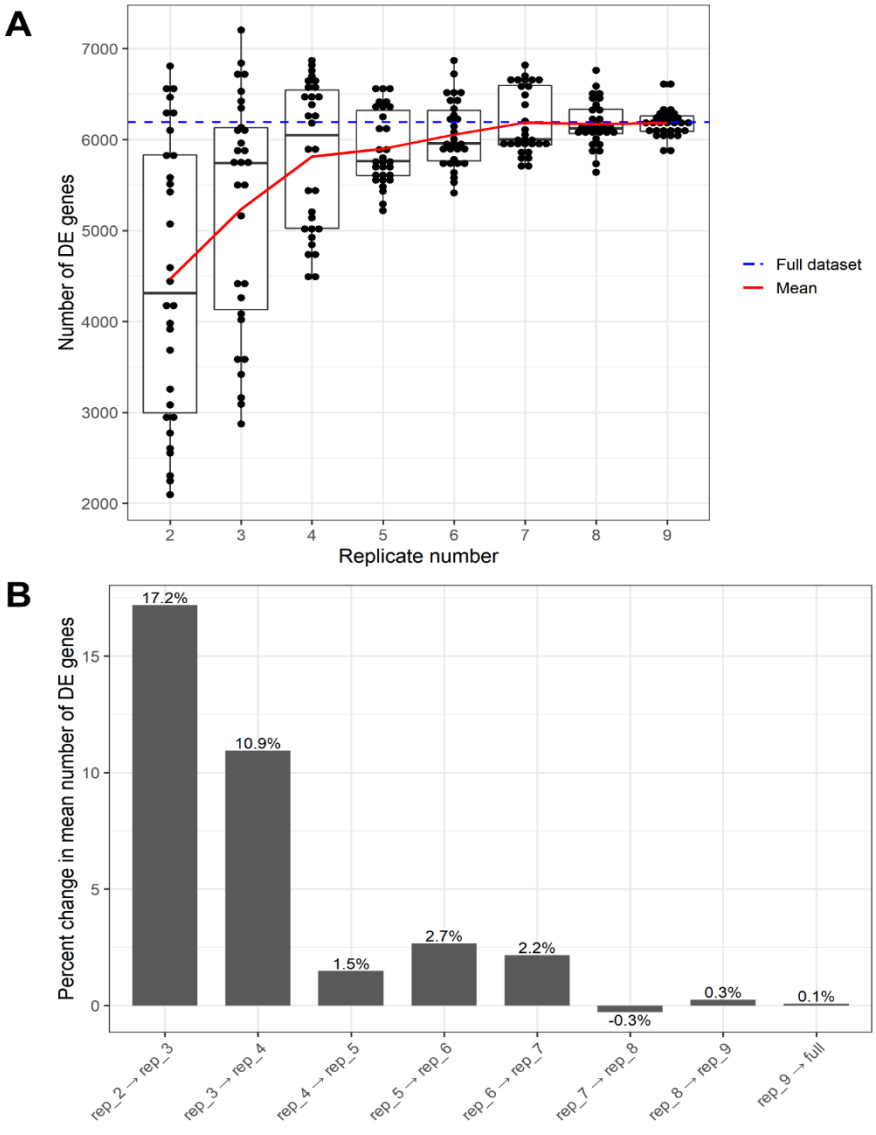

Figure 5.3 *ERSSA* with 10 replicate GTEx dataset – part 1
Plots generated by running *ERSSA* with 10 replicate GTEx dataset comparing human muscle and heart tissues. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Dots represent the number of DE genes discovered in each subsample. Boxplot summarizes the results. Red and blue lines represent the average and full discovery, respectively. **B.** Percent change in average DE gene discovery as sample size increases.

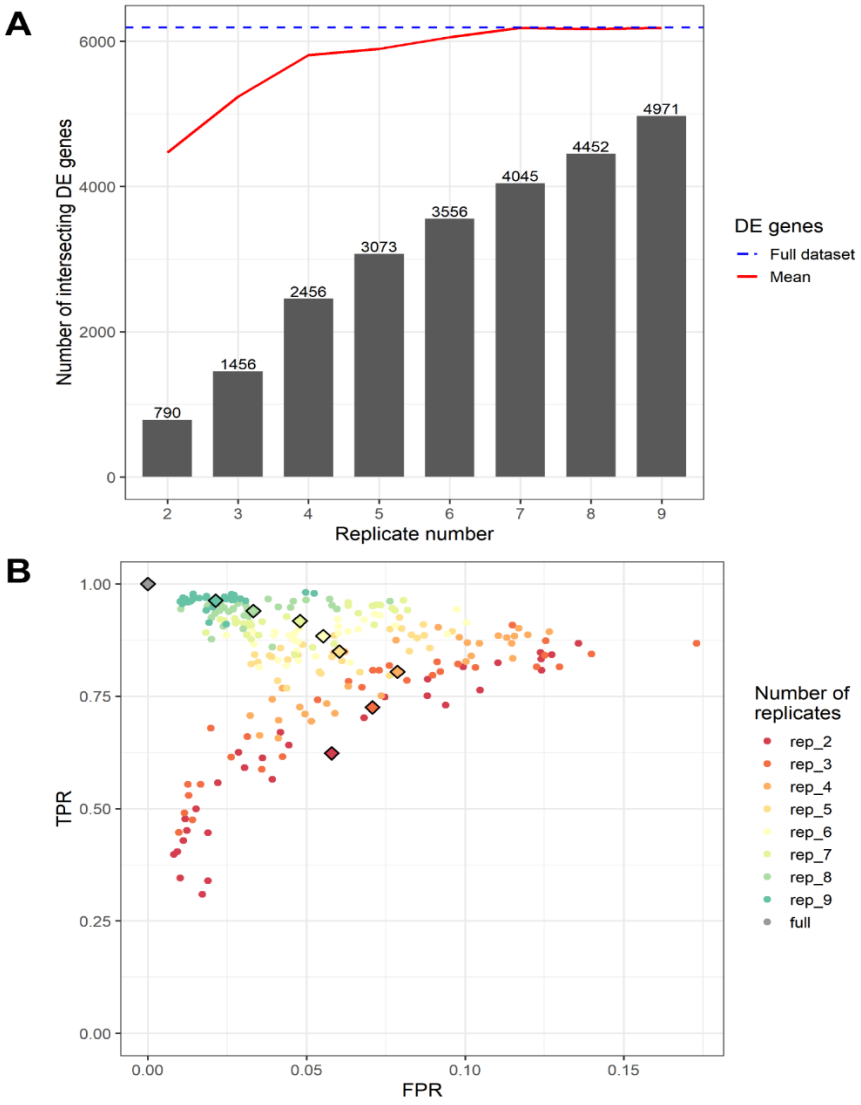

Figure 5.4 *ERSSA* with 10 replicate GTEx dataset – part 2
Plots generated by running *ERSSA* with 10 replicate GTEx dataset comparing human muscle and heart tissues. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Bar plot of the number of DE genes found in all subsamples at each replicate level. Red and blue lines represent the average and full discovery, respectively. **B.** Dots indicate FPR and TPR of each subsample. Diamonds represent the average rates at each replicate level. DE discovery with the full dataset is set as the ground truth for the calculations.

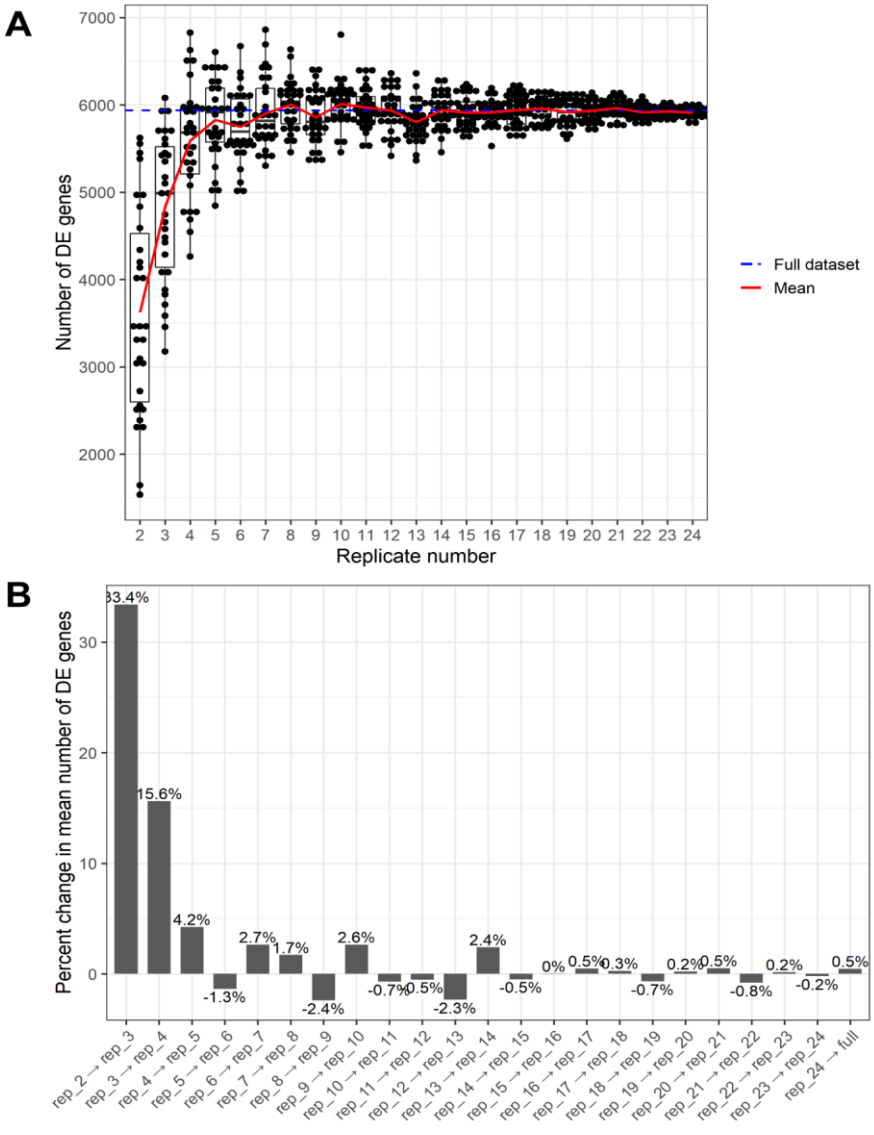

Figure 5.5 *ERSSA* with 25 replicate GTEx dataset – part 1
Plots generated by running *ERSSA* with 25 replicate GTEx dataset comparing human muscle and heart tissues. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Dots represent the number of DE genes discovered in each subsample. Boxplot summarizes the results. Red and blue lines represent the average and full discovery, respectively. **B.** Percent change in average DE gene discovery as sample size increases.

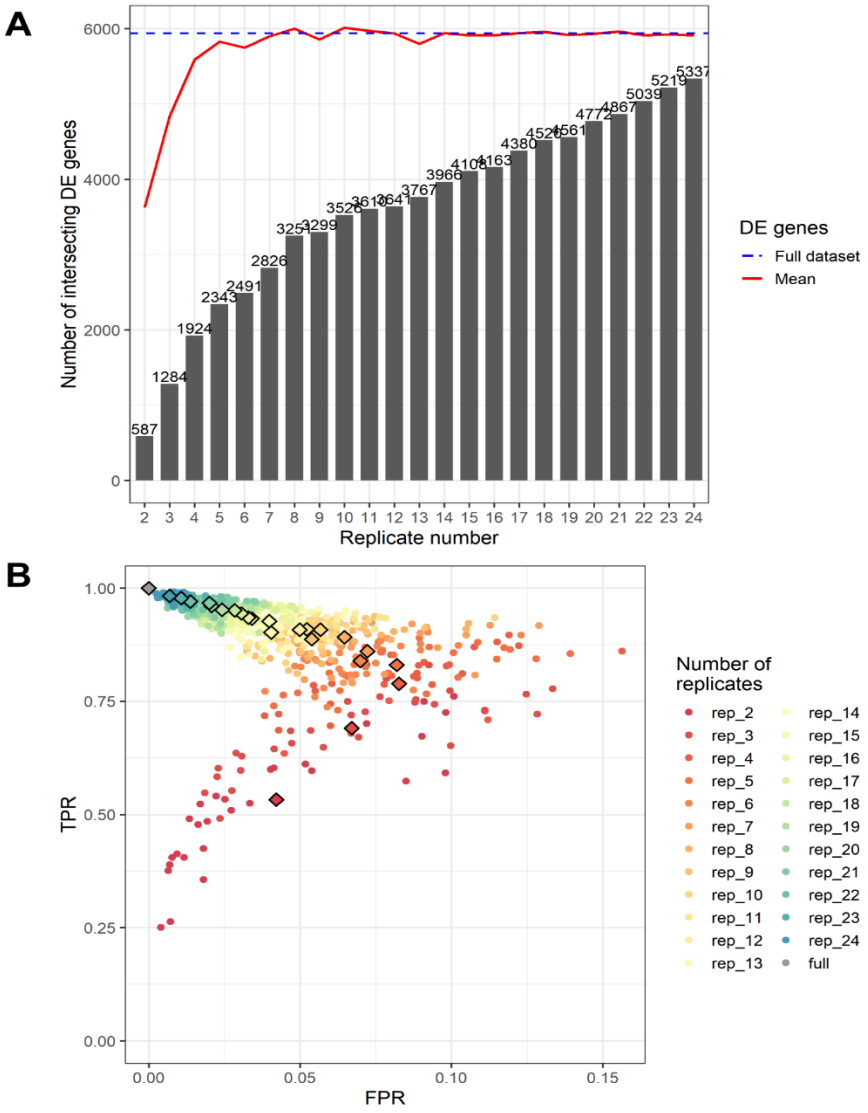

Figure 5.6 *ERSSA* with 25 replicate GTEx dataset – part 2
Plots generated by running *ERSSA* with 25 replicate GTEx dataset comparing human muscle and heart tissues. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Bar plot of the number of DE genes found in all subsamples at each replicate level. Red and blue lines represent the average and full discovery, respectively. **B.** Dots indicate FPR and TPR of each subsample. Diamonds represent the average rates at each replicate level. DE discovery with the full dataset is set as the ground truth for the calculations.

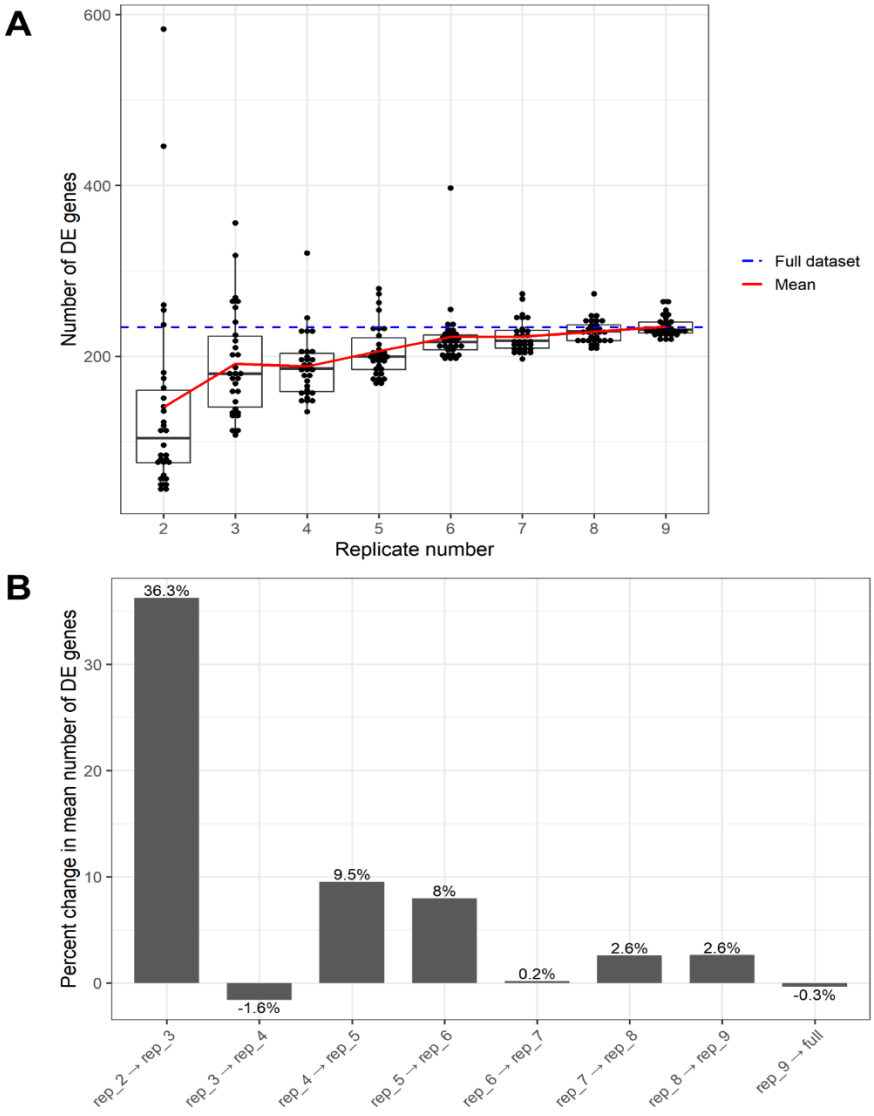

Figure 5.7 *ERSSA* with Bottomly dataset – part 1
Plots generated by running *ERSSA* with Bottomly dataset comparing brain tissues from two mouse strains. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Dots represent the number of DE genes discovered in each subsample. Boxplot summarizes the results. Red and blue lines represent the average and full discovery, respectively. **B.** Percent change in average DE gene discovery as sample size increases.

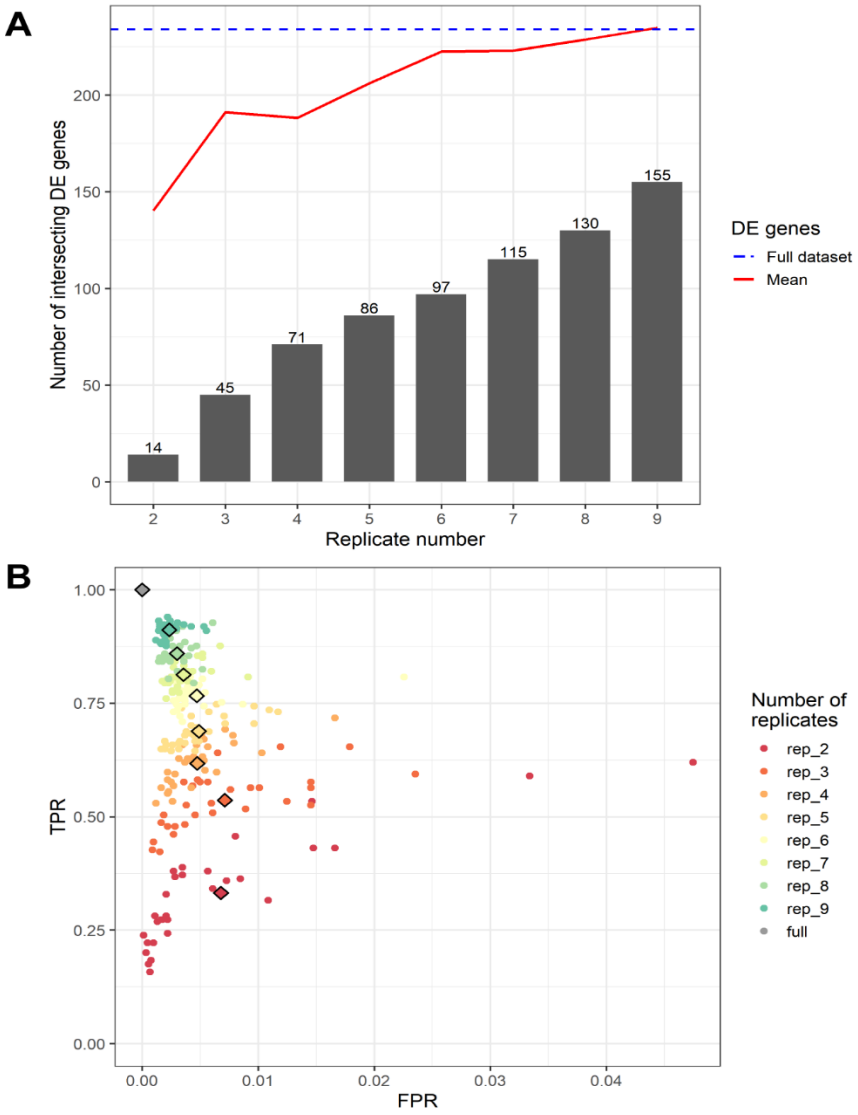

Figure 5.8 *ERSSA* with Bottomly dataset – part 2
Plots generated by running *ERSSA* with Bottomly dataset comparing brain tissues from two mouse strains. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. **A.** Bar plot of the number of DE genes found in all subsamples at each replicate level. Red and blue lines represent the average and full discovery, respectively. **B.** Dots indicate FPR and TPR of each subsample. Diamonds represent the average rates at each replicate level. DE discovery with the full dataset is set as the ground truth for the calculations.

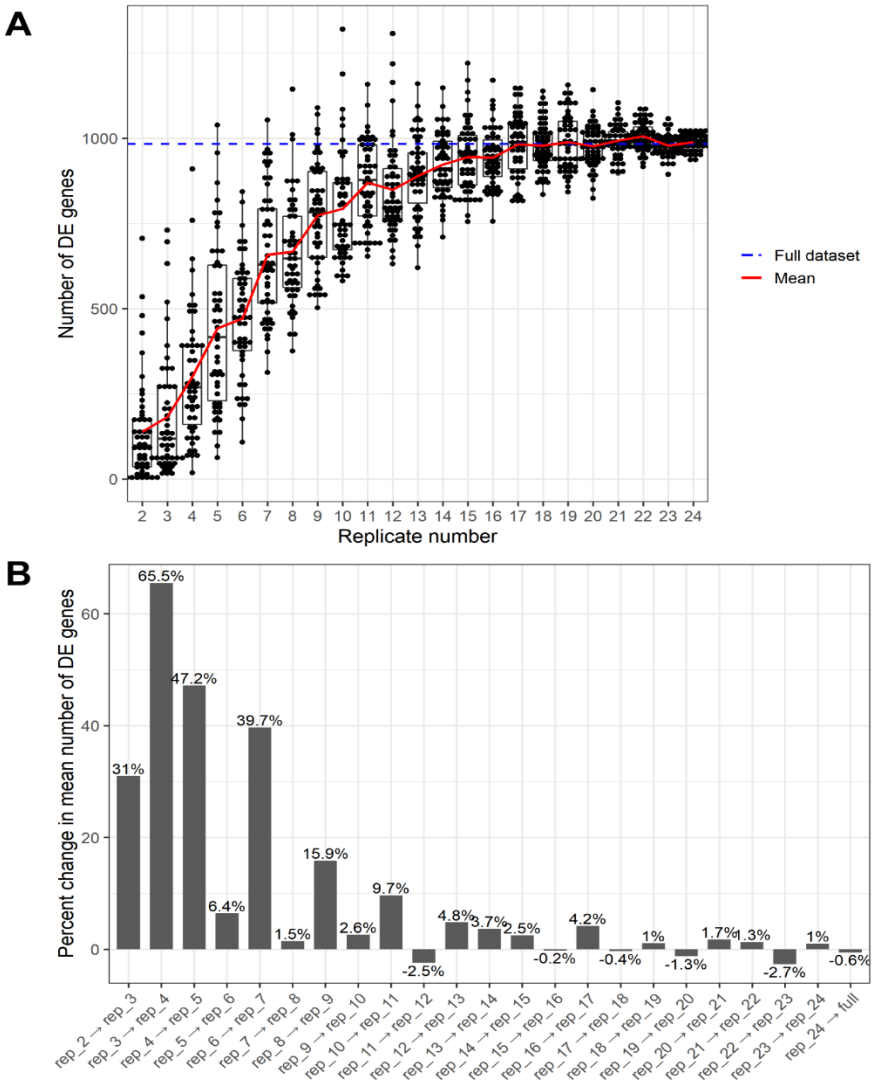

Figure 5.9 *ERSSA* with MP dataset – part 1
Plots generated by running *ERSSA* with 25 replicate MP dataset comparing lymphoblastoid cell lines derived from Europeans and Nigerian individuals. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 50 subsamples at each replicate level. **A.** Dots represent the number of DE genes discovered in each subsample. Boxplot summarizes the results. Red and blue lines represent the average and full discovery, respectively. **B.** Percent change in average DE gene discovery as sample size increases.

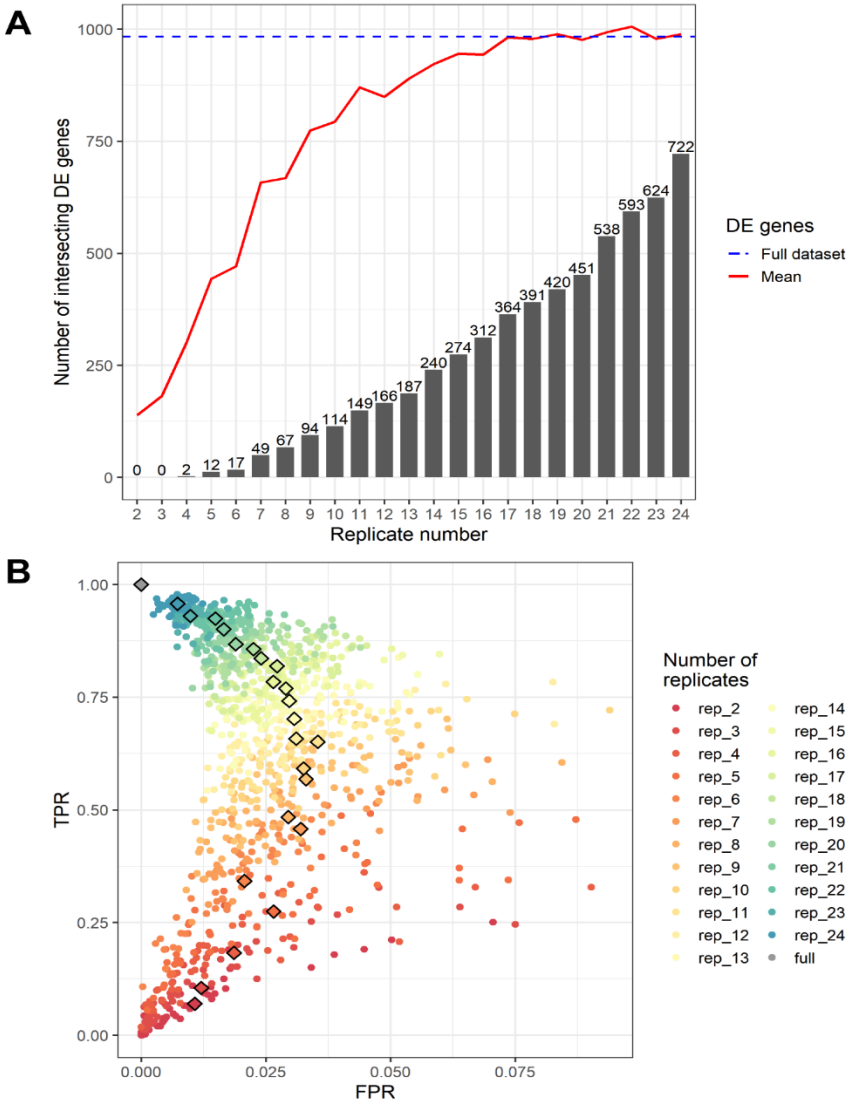

Figure 5.10 *ERSSA* with MP dataset – part 2
Plots generated by running *ERSSA* with 25 replicate MP dataset comparing lymphoblastoid cell lines derived from Europeans and Nigerian individuals. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 50 subsamples at each replicate level. **A.** Bar plot of the number of DE genes found in all subsamples at each replicate level. Red and blue lines represent the average and full discovery, respectively. **B.** Dots indicate FPR and TPR of each subsample. Diamonds represent the average rates at each replicate level. DE discovery with the full dataset is set as the ground truth for the calculations.

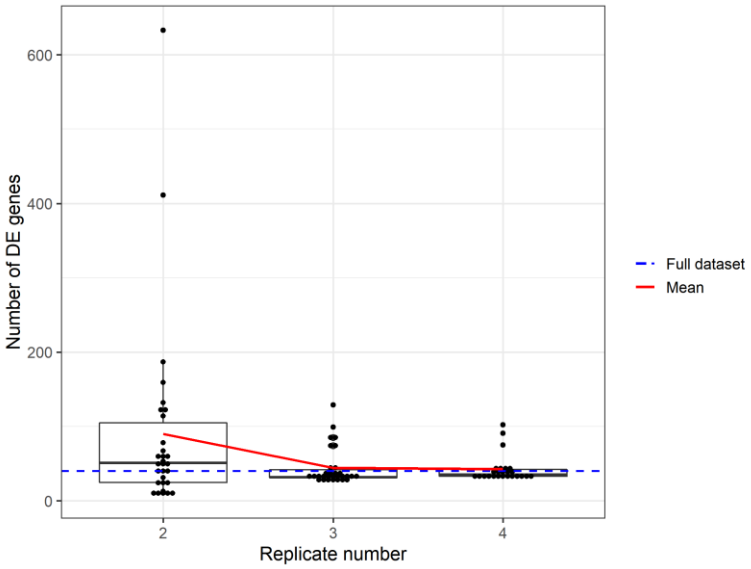

Figure 5.11 *ERSSA* with Fossum dataset using absolute log2-fold-change cutoff of 1.0
Plot generated by running *ERSSA* with Fossum dataset comparing cell line treated with transcription factor knockdown or control siRNA. Default settings used for DE definition (FDR<0.05 and |log2-fold-change|>1). Up to 30 subsamples at each replicate level. Very few DE genes are discovered with the full dataset using the above cutoffs. Dots represent the number of DE genes discovered in each subsample. Boxplot summarizes the results. Red and blue lines represent the average and full discovery, respectively.

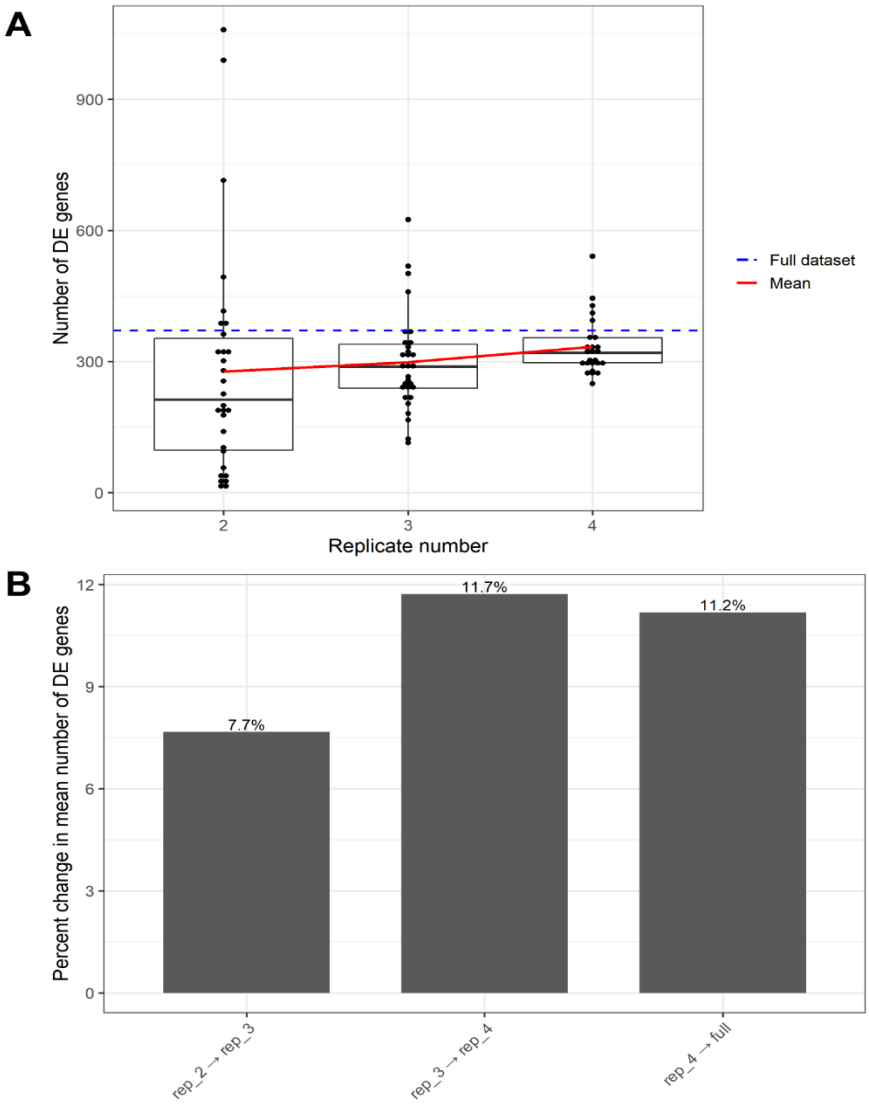

Figure 5.12 *ERSSA* with Fossum dataset – part 1
Plot generated by running *ERSSA* with Fossum dataset comparing cell line treated with transcription factor knockdown or control siRNA. DE gene defined to have FDR<0.05 and |log2-fold-change|>0.5. Up to 30 subsamples at each replicate level. **A.** Dots represent the number of DE genes discovered in each subsample. Boxplot summarizes the results. Red and blue lines represent the average and full discovery, respectively. **B.** Percent change in average DE gene discovery as sample size increases.

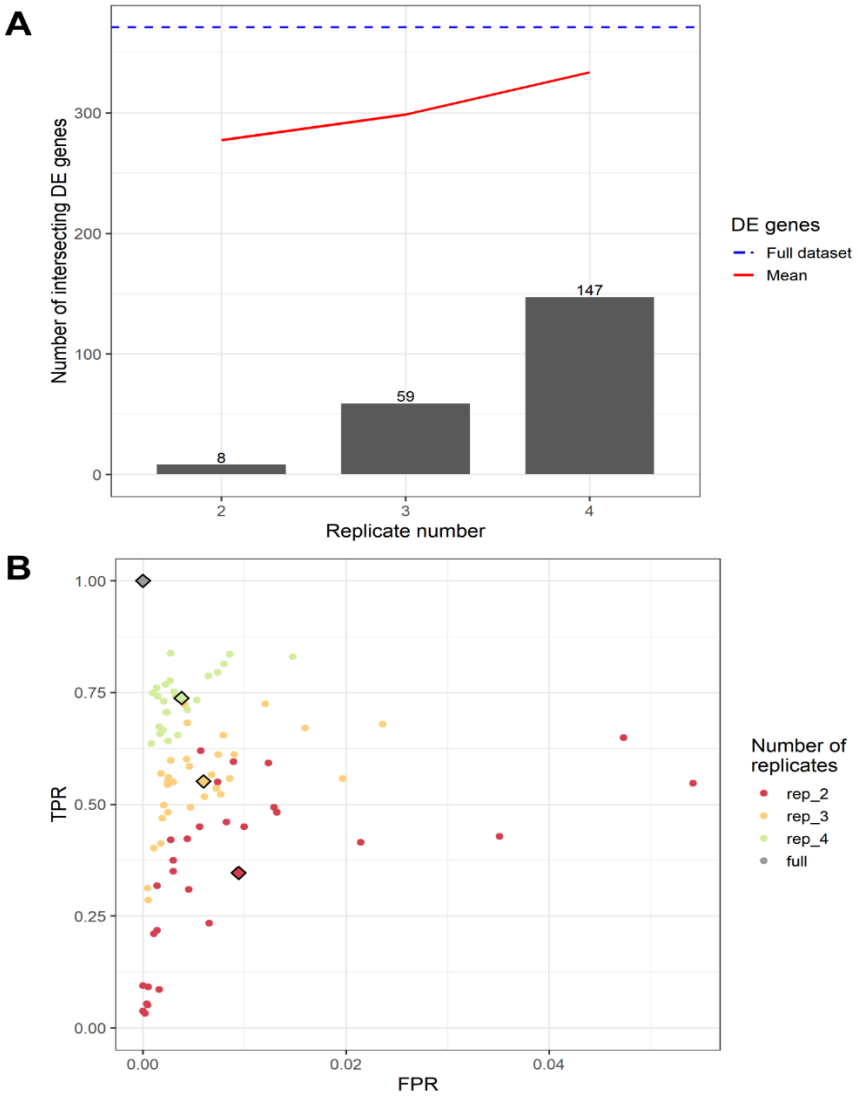

Figure 5.13 *ERSSA* with Fossum dataset – part 2
Plot generated by running *ERSSA* with Fossum dataset comparing cell line treated with transcription factor knockdown or control siRNA. DE gene defined to have FDR<0.05 and |log2-fold-change|>0.5. Up to 30 subsamples at each replicate level. **A.** Bar plot of the number of DE genes found in all subsamples at each replicate level. Red and blue lines represent the average and full discovery, respectively. **B.** Dots indicate FPR and TPR of each subsample. Diamonds represent the average rates at each replicate level. DE discovery with the full dataset is set as the ground truth for the calculations.

Sample ID	Condition	N=25	N=10	N=5
SRR598148	heart	x	x	x
SRR598509	heart	x	x	x
SRR598589	heart	x	x	x
SRR599025	heart	x	x	x
SRR599086	heart	x	x	x
SRR599249	heart	x	x	
SRR599380	heart	x	x	
SRR600474	heart	x	x	
SRR600829	heart	x	x	
SRR600852	heart	x	x	
SRR600924	heart	x		
SRR601239	heart	x		
SRR601613	heart	x		
SRR601645	heart	x		
SRR601868	heart	x		
SRR601986	heart	x		
SRR602106	heart	x		
SRR602437	heart	x		
SRR602461	heart	x		
SRR603449	heart	x		
SRR603918	heart	x		
SRR603968	heart	x		
SRR604122	heart	x		
SRR604174	heart	x		
SRR604206	heart	x		

Table 5.1 GTEx heart samples tested with *ERSSA*

25 GTEx human heart samples included in the analysis. A portion of the samples were used to simulate experiments with 5 and 10 replicates. Samples with a cross indicates inclusion in the respective analysis (example, sample SRR598148 is included in all three analyses while sample SRR600924 is only used in the N=25 analysis).

Sample ID	Condition	N=25	N=10	N=5
SRR598044	muscle	x	x	x
SRR598452	muscle	x	x	x
SRR600656	muscle	x	x	x
SRR600981	muscle	x	x	x
SRR601387	muscle	x	x	x
SRR601671	muscle	x	x	
SRR601695	muscle	x	x	
SRR601815	muscle	x	x	
SRR602010	muscle	x	x	
SRR603116	muscle	x	x	
SRR603164	muscle	x		
SRR603236	muscle	x		
SRR605101	muscle	x		
SRR606974	muscle	x		
SRR607067	muscle	x		
SRR607117	muscle	x		
SRR608264	muscle	x		
SRR612215	muscle	x		
SRR612227	muscle	x		
SRR612395	muscle	x		
SRR612539	muscle	x		
SRR612635	muscle	x		
SRR612683	muscle	x		
SRR612695	muscle	x		
SRR612767	muscle	x		

Table 5.2 GTEx muscle samples tested with *ERSSA*

25 GTEx human muscle samples included in the analysis. A portion of the samples were used to simulate experiments with 5 and 10 replicates. Samples with a cross indicates inclusion in the respective analysis (example, sample SRR598044 is included in all three analysis while sample SRR603164 is only used in the N=25 analysis).

Number of Replicates [a]	N=5, Union of all DE genes[b]	N=5, Consensus DE genes [%][c]	N=10, Union of all DE genes[b]	N=10, Consensus DE genes [%][c]	N=25, Union of all DE genes[b]	N=25, Consensus DE genes [%][c]
n = 2	9102	6.62	10545	7.49	10315	5.69
n = 3	8789	18.02	10924	13.33	11370	11.29
n = 4	8108	39.38	10630	23.10	10984	17.52
n = 5			9751	31.51	10873	21.55
n = 6			9343	38.06	10234	24.34
n = 7			8750	46.23	10130	27.90
n = 8			8119	54.83	9369	34.70
n = 9			7516	66.14	9002	36.65
n = 10					9018	39.10
n = 11					8765	41.19
n = 12					8573	42.47
n = 13					8178	46.06
n = 14					8165	48.57
n = 15					7855	52.30
n = 16					7790	53.44
n = 17					7636	57.36
n = 18					7499	60.27
n = 19					7259	62.83
n = 20					7103	67.18
n = 21					7000	69.53
n = 22					6829	73.79
n = 23					6613	78.92
n = 24					6440	82.87

Table 5.3 Effect of the number of replicates on the fidelity of DE gene identification: GTEx human muscle and human heart dataset

Analysis repeated for all three GTEx sample size (*N*=5, 10 and 25). [a]A random selection of up to 30 unique subsamples were analyzed at each replicate level. [b]The number of putative DE genes that were found in at least one of the subsample analyses. [c]The fraction of those genes that were identified in every one of the subsample analyses.

Sample ID	Condition
SRX033480	C57BL/6J
SRX033488	C57BL/6J
SRX033481	C57BL/6J
SRX033489	C57BL/6J
SRX033482	C57BL/6J
SRX033490	C57BL/6J
SRX033483	C57BL/6J
SRX033476	C57BL/6J
SRX033478	C57BL/6J
SRX033479	C57BL/6J
SRX033472	DBA/2J
SRX033473	DBA/2J
SRX033474	DBA/2J
SRX033475	DBA/2J
SRX033491	DBA/2J
SRX033484	DBA/2J
SRX033492	DBA/2J
SRX033485	DBA/2J
SRX033493	DBA/2J
SRX033486	DBA/2J
SRX033494	DBA/2J

Table 5.4 Bottomly dataset
The Bottomly dataset contains 10 brain tissue samples from C57BL/6J mouse strain and 11 brain tissue samples from DBA/2J strain. All samples are used in *ERSSA*.

Number of Replicates[a]	Union of all DE genes[b]	Consensus DE genes [%][c]
n = 2	1099	1.27
n = 3	1046	4.30
n = 4	771	9.21
n = 5	720	11.94
n = 6	677	14.33
n = 7	512	22.46
n = 8	434	29.95
n = 9	373	41.55

Table 5.5 Effect of the number of replicates on the fidelity of DE gene identification: Bottomly dataset

[a]A random selection of up to 30 unique subsamples were analyzed at each replicate level.
[b]The number of putative DE genes that were found in at least one of the subsample analyses.
[c]The fraction of those genes that were identified in every one of the subsample analyses.

Sample ID	Condition	Sample ID	Condition
NA06985	European	NA18486	Nigerian
NA06986	European	NA18498	Nigerian
NA06994	European	NA18499	Nigerian
NA07000	European	NA18501	Nigerian
NA07037	European	NA18502	Nigerian
NA07051	European	NA18504	Nigerian
NA07346	European	NA18505	Nigerian
NA07347	European	NA18507	Nigerian
NA07357	European	NA18508	Nigerian
NA10847	European	NA18510	Nigerian
NA10851	European	NA18511	Nigerian
NA11829	European	NA18516	Nigerian
NA11830	European	NA18517	Nigerian
NA11831	European	NA18519	Nigerian
NA11832	European	NA18520	Nigerian
NA11840	European	NA18522	Nigerian
NA11881	European	NA18523	Nigerian
NA11894	European	NA18852	Nigerian
NA11918	European	NA18853	Nigerian
NA11920	European	NA18855	Nigerian
NA11931	European	NA18856	Nigerian
NA11992	European	NA18858	Nigerian
NA11993	European	NA18861	Nigerian
NA11994	European	NA18862	Nigerian
NA11995	European	NA18870	Nigerian

Table 5.6 MP dataset

25 replicates from each group were randomly selected from the original MP dataset for the purpose of *ERSSA* analysis.

Number of Replicates[a]	Union of all DE genes[b]	Consensus DE genes [%][c]
n = 2	2085	0.00
n = 3	2399	0.00
n = 4	2936	0.07
n = 5	3265	0.37
n = 6	2892	0.59
n = 7	3194	1.53
n = 8	3122	2.15
n = 9	3018	3.11
n = 10	3056	3.73
n = 11	2972	5.01
n = 12	2819	5.89
n = 13	2700	6.93
n = 14	2477	9.69
n = 15	2428	11.29
n = 16	2269	13.75
n = 17	2202	16.53
n = 18	2073	18.86
n = 19	1985	21.16
n = 20	1835	24.58
n = 21	1707	31.52
n = 22	1608	36.88
n = 23	1471	42.42
n = 24	1364	52.93

Table 5.7 Effect of the number of replicates on the fidelity of DE gene identification: MP dataset

[a]A random selection of up to 50 unique subsamples were analyzed at each replicate level.
[b]The number of putative DE genes that were found in at least one of the subsample analyses.
[c]The fraction of those genes that were identified in every one of the subsample analyses.

Sample ID	Condition
SRR1655002	Control_siRNA
SRR1655003	Control_siRNA
SRR1655004	Control_siRNA
SRR1655005	Control_siRNA
SRR1655006	Control_siRNA
SRR1655007	EHF_siRNA
SRR1655008	EHF_siRNA
SRR1655009	EHF_siRNA
SRR1655010	EHF_siRNA
SRR1655011	EHF_siRNA

Table 5.8 Fossum dataset
The Fossum dataset contains 5 replicates each of cell lines exposed to either control or EHF knockdown siRNA.

Number of Replicates[a]	Union of all DE genes[b]	Consensus DE genes [%][c]
n = 2	2312	0.35
n = 3	1290	4.57
n = 4	837	17.56

Table 5.9 Effect of the number of replicates on the fidelity of DE gene identification: Fossum dataset
[a]A random selection of up to 30 unique subsamples were analyzed at each replicate level.
[b]The number of putative DE genes that were found in at least one of the subsample analyses.
[c]The fraction of those genes that were identified in every one of the subsample analyses.

5.7 References

1. Ching, T., Huang, S. & Garmire, L. X. Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* (2014). doi:10.1261/rna.046011.114
2. Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301–304 (2014).
3. Hoskins, S. P., Shyr, D. & Shyr, Y. Sample Size Calculation for Differential Expression Analysis of RNA-Seq Data. in *Frontiers of Biostatistical Methods and Applications in Clinical Oncology* 359–379 (Springer, Singapore, 2017). doi:10.1007/978-981-10-0126-0_22
4. Poplawski, A. & Binder, H. Feasibility of sample size calculation for RNA-seq studies. *Brief. Bioinform.* (2017). doi:10.1093/bib/bbw144
5. Shao, Z. ERSSA: Empirical RNA-seq Sample Size Analysis. *Bioconductor* Available at: http://bioconductor.org/packages/ERSSA.
6. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
7. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
8. Shao, Z. ERSSA Package Introduction. Available at: http://bioconductor.org/packages/3.8/bioc/vignettes/ERSSA/inst/doc/ERSSA.html. (Accessed: 9th October 2018)
9. Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
10. Bottomly, D. *et al.* Evaluating Gene Expression in C57BL/6J and DBA/2J Mouse Striatum Using RNA-Seq and Microarrays. *PLOS ONE* **6**, e17820 (2011).
11. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, (2010).
12. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
13. Fossum, S. L. *et al.* Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells. *Nucleic Acids Res.* **42**, 13588–13598 (2014).
14. Frazee, A. C., Langmead, B. & Leek, J. T. ReCount: A multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics* **12**, 449 (2011).
15. Collado-Torres, L. *et al.* Reproducible RNA-seq analysis using recount2. *Nat. Biotechnol.* **35**, 319–321 (2017).
16. Collado-Torres, L., Nellore, A. & Jaffe, A. E. recount workflow: Accessing over 70,000 human RNA-seq samples with Bioconductor. *F1000Research* **6**, 1558 (2017).

17. Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8**, 1765–1786 (2013).
18. Levin, C. Best bioinformatics software for RNA-seq quantification and differential expression. *omictools* Available at: https://omictools.com/blog/your-top-3-rna-seq-quantification-and-differential-expression-tools. (Accessed: 17th September 2018)
19. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2009).
20. Busby, M. A., Stewart, C., Miller, C. A., Grzeda, K. R. & Marth, G. T. Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. *Bioinforma. Oxf. Engl.* **29**, 656–657 (2013).
21. Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A. & Kocher, J.-P. Calculating sample size estimates for RNA sequencing data. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **20**, 970–978 (2013).
22. Bi, R. & Liu, P. Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-sequencing experiments. *BMC Bioinformatics* **17**, 146 (2016).
23. Vieth, B., Ziegenhain, C., Parekh, S., Enard, W. & Hellmann, I. powsimR: power analysis for bulk and single cell RNA-seq experiments. *Bioinformatics* (2017). doi:10.1093/bioinformatics/btx435
24. BiocParallel: Bioconductor facilities for parallel evaluation. *Bioconductor* (2018). Available at: http://bioconductor.org/packages/BiocParallel/. (Accessed: 19th September 2018)

Chapter 6. Reinforcing Polylactide (PLA) with Tungsten Disulfide (WS2) nanotubes to enable thinner, radio-opaque bioresorbable vascular scaffolds

6.1 Introduction

The treatment of coronary heart disease (CHD), one of the leading causes of death in the world[1,2], has seen significant breakthroughs due to improvements in percutaneous cardiovascular intervention. Drug-eluting metal stents (DES), the current standard of care for CHD, restore blood flow through the occluded artery by physically supporting the artery at the site of the lesion. Despite their success, DES are made from materials (e.g. Co-Cr alloys[3]) that are not biodegradable and consequently, lead to undesirable side effects. Patients implanted with metal stents (>1M in the US alone in 2008[4]) suffer from angina, due to restricted arterial vasomotion, and are at risk of developing late stent thrombosis (LST), the most dreaded complication associated with stents[5,6]. A promising successor to DES is a bioresorbable vascular scaffold (BVS), which is made entirely out of a biodegradable polymer such as poly L-lactide (PLLA)[7,8]. Unlike DES, the first clinically-approved BVS (CE Mark in 2011[5] and FDA-approval in 2016[9]) is a transient entity in the body; it supports the occluded artery for the requisite 3-6 months but is completely resorbed in 2-3 years[10].

Clinical reports on the five-year follow-up to the first-in-man trials for BVS indicated that they restore arterial vasomotion and eliminate the risk of LST[11–13]. However, subsequent clinical trials reported an increase in thrombosis for BVS compared to metal stents within

the first year[14]. Clinicians speculate that the greater thickness of BVSs (~150μm) relative to metal stents (~80μm) perturbs the flow of blood which contributes towards the risk of thrombosis[12]. Surgeons also find it challenging to implant a BVS under X-ray guidance as polymers are virtually transparent to X-rays. This difficulty in "seeing" the BVS during implantation can result in malapposed scaffolds at the site of the lesion, which in turn contributes towards thrombosis. Thus, physicians advocate for two main improvements in BVS to foster wider adoption and to serve a broader spectrum of patients: (1) reduce the thickness of BVS to treat smaller vessels and mitigate the risk of thrombosis, and (2) enhance the radio-opacity of BVS to improve visualization with X-rays.

The key to a thinner ~80μm BVS is to confer radial strength comparable to the clinically-approved 150μm BVS. PLLA can be made stronger via blending with other polymers[15,16] or copolymerization[17–19], but these methods do not provide any increase in radio-opacity and can lead to premature loss of radial strength due to accelerated degradation in the body[20]. Therefore, reinforcing PLLA with inorganic, radio-opaque nanoparticles seems to be a viable solution for tackling the dual challenge of a thinner, radio-opaque vascular scaffold. Widely used nanoparticles such as carbon nanotubes and graphene nanosheets can increase the strength of the polymer matrix[21], but they disperse poorly without functionalization[22], provide no increase in radio-opacity, and are toxic to mammalian cells[23]. Nanocomposites made with Group 3 (Mg, Zn, and Fe) and Group 4 (Ti) elements have been unsuccessful due to the lack of sufficient radio-opacity. Group 6 elements (W, Au, and Pt) are promising candidates as they provide radio-opacity comparable to an ~80μm-thick Co-Cr stent at a

relatively low volume fraction of ~6%. Of these three elements, we select Tungsten Disulfide (WS_2) nanotubes (WSNTs) as they readily disperse in PLLA without any surface modification[24] and show promising biocompatibility *in vitro*[25–27]. We select WSNTs over other WS_2 nanoparticles (e.g spheres or fullerene-like) as their high-aspect ratio (70–200 nm diameter, 2–3 μm length) will favor preferential orientation along the hoop-direction of the BVS during processing, which can enhance radial strength.

The clinically-approved BVS is made from a predominantly amorphous PLLA preform that is processed through a sequence of tube expansion, laser-cutting and crimping. Tube expansion subjects the PLLA preform to strains in excess of 400% on the order of seconds. This rapid deformation transforms the initially amorphous preform into a highly oriented, semicrystalline tube[28]. The expanded tube is subsequently laser-cut to create an "as-cut" scaffold with a lattice network of struts that permits crimping onto a balloon catheter. Crimping is performed near the glass transition temperature of PLLA and confers a unique morphology that facilitates deployment in the artery[29] and provides lasting radial strength for months afterword[30]. Nanoparticles are known to alter morphology development during processing, particularly in semicrystalline polymers[21,22], motivating the present study on the biocompatibility of WSNT-reinforced polylactide (PLA, <2% D-content) and the effects of WSNTs on the semicrystalline morphology of PLA under the influence of flow. Guided by prior literature showing that 0.1 wt% of WSNTs boost the compressive strength of polypropylenefumarate (PPF) by over 50%[31], we examine up to 0.1 wt% of WSNTs in PLA. We assess the biocompatibility of bare WSNTs and PLA-WSNT films with appropriate

controls via microscopy and cytotoxicity assays in cells relevant to the human vasculature. In view of the importance of kinetically-controlled morphology in the clinically-approved BVS[29,30], we study the development of PLA structure in real time under elongational and shear flow. In contrast to the scant literature on injection-molded PLA that only probes the final state of the material, we perform *in situ* measurements that reveal aspects of the PLA morphology that are only observed for short intervals during flow (e.g. inception of thread-like precursors that template oriented crystals). The flow-induced crystallization experiments described in this report place an emphasis on the shear stress, which correlates with the degree of orientation of chain segments in the melt. We complement the *in situ* measurements with *ex situ* microscopy and X-ray scattering to provide deeper insight on the impact of WSNTs on the PLA morphology both during and post flow.

6.2 Materials and Methods

Preparation of nanoparticles

Tungsten Disulfide (WS_2) nanotubes (WSNTs; 70-200 nm diameter, 2-3 μm length) (**Figure S1A**) were purchased from NanoMaterials, Israel. Before use, the WSNTs were sonicated and centrifuged to eliminate agglomerates and small impurities (e.g. broken nanotubes) using the following procedure. First, 200 mg of WSNTs was dispersed in 200 mL of isopropyl alcohol; the resulting solution was then sonicated in a flask for 2 hours. Second, the sonicated solution was poured into falcon tubes (leaving behind the sediment at the bottom of the flask) and centrifuged at 1500 rpm for 15 minutes. Third, the supernatant was discarded and falcon

tubes containing the residue were immersed in a water bath at 50-60°C for a few days until the WSNTs were completely dry and ready to use. Zinc Oxide (ZnO) nanoparticles (catalogue #: 544906, Sigma-Aldrich) (**Figure S1B**) were used as a negative control in cytotoxicity experiments. For cell culture, the nanoparticles (WSNTs and ZnO) were sterilized by UV light overnight before they were solubilized in media using a combination of sonication and vortexing. Residual, non-solubilized aggregates were filtered out using a cell strainer with a 40 μm pore size. The nanoparticle-media mixture was then added to cell culture wells to reach the desired final treatment concentration.

Preparation of nanocomposite films

This study uses polylactide (PLA) with ~2% D-content[32,33] and M_w ~ 125 kg/mol (values in **Figure S2**; gel-permeation chromatograms in **Figure S3**) purchased from NatureWorks, USA (Grade 4032D). To limit degradation of PLA at high temperatures (>150°C), PLA pellets are first dried under vacuum at 80°C for 1 day and subsequently at 40°C for an additional 4 days to reduce moisture. The vacuum-dried polymer is then dissolved in chloroform at 60°C for 2 hours. In parallel, a separate solution of WSNTs (0.05 or 0.1wt%) in chloroform is sonicated for ~30 mins to break up nanotube agglomerates. The solution of WSNTs is then added to the polymer solution and stirred for an additional 1 hour; the resulting solution is then cast into glass petri dishes and allowed to dry inside a fume-hood for ~2 days; the as-cast films are placed in an oven and dried for ~4 days under vacuum at 80°C to remove residual solvent. The films are thereafter maintained at 40°C under vacuum.

SEM of nanoparticles

Scanning electron micrographs of WSNTs and ZnO nanoparticles (**Figure S1**) were acquired at Caltech on a ZEISS 1550VP Field Emission SEM with an electron accelerating voltage of 10kV. The SEM samples were prepared using the following protocol: (1) disperse 7.2 mg of nanoparticles (WSNTs or ZnO) in 12 mL of chloroform, (2) sonicate the dispersion for 20 minutes to break-up agglomerates, and (3) drop-cast the sonicated solution (~0.5 mL) on a silicon wafer for imaging.

Cell culture

Pooled, Normal, Primary Umbilical Vein Endothelial Cells (HUVECs, ATCC PCS-100-013) and Normal Human Primary Aortic Smooth Muscle Cells (HASMCs, ATCC PCS-100-012) were obtained from the American Type Culture Collection (ATCC). HUVECs were grown in Vascular Cell Basal Medium with Endothelial Cell Growth Kit-BBE additive (ATCC). HASMCs were grown in Vascular Cell Basal Medium with Vascular Smooth Muscle Cell Growth Kit additive (ATCC). P5 to P7 HUVECs and P4 to P6 HASMCs were used in all cell culture experiments. All cell cultures were grown in a humidified incubator at 37˚C with 5% carbon dioxide to control media pH.

Preparation of PLA, PLA-WSNT and PLA-ZnO disks for cell culture

Disks were prepared by first cutting out circles (diameter: 22 mm) from solvent-cast PLA, PLA-WSNT (0.1 wt%), and PLA-ZnO films and then punching out their center (inner diameter: 6 mm) to facilitate transport of media (**Figure S4A-B**). The disks were then

sterilized through stringent washes in a 70% Ethanol, 30% water solution followed by thorough air drying. The dried disks were then inserted into 12-well cell culture plates at the start of the treatment (**Figure S4C**). At end of the treatment, the disks were carefully removed to facilitate assay reagent diffusion prior to the WST-1 assay and Live/Dead staining.

WST-1 assay

The WST-1 cell proliferation reagent (Roche) was first added to the cell culture media at the recommended working concentration; the reagent-media solution was subsequently incubated for four hours at 37°C. After incubation, 100 µL of the solution was transferred to each well of the 96-well plate for plate-reading using a Flexstation 3 microplate reader (440 nm and 690 nm). The baseline absorption from the background and media was subtracted to isolate the contribution of the cells. All test wells were then normalized as a percentage of the mean control well value except for the disk overlay study, in which normalization was done as a percentage of the mean PLA well value. In the bare nanoparticle concentration series and time series experiments, statistical tests were performed between control and treated cells. In the disk overlay experiment, statistical tests were performed between cells exposed to PLA films and other conditions. A 10x Lysis Buffer (Thermo Scientific Pierce) was used as a lethal control (cells were treated for 15 mins prior to the WST-1 assay). For the time series experiments, each disk was incubated with 1 mL of the appropriate growth media for 24 hours in a humidified incubator maintained at 37°C. Cells were then exposed to this pre-conditioned media (growth media that was in contact with the disks) and their viability was assessed via WST-1 at three different time points (24, 48, and 72 hours).

Live/Dead staining and phase-contrast microscopy

The LIVE/DEAD Viability/Cytotoxicity Kit (Invitrogen) was used to simultaneously capture the status of both dead and live cells. At start of the assay, cells were washed once with Dulbecco's Phosphate-Buffered Saline (DPBS) before incubation for 30 minutes at 37˚C in the media-dye mixture. At end of incubation, cells were washed with DPBS and imaged on a Zeiss Axiovert 25CFL microscope. The FIJI software package was used to merge fluorescent signals into one image with red indicating dead cells and green indicating live cells. Phase contrast images were captured on the same microscopy system with bright-field illumination (see **Figure S5** for phase contrast micrographs of the cells before and after the wash). A 10x Lysis Buffer (Thermo Scientific Pierce) was used as the lethal control and was added to wells for 15 minutes before Live/Dead staining.

TEM of cells cultured with WSNT

Cells were grown in 100 mm cell culture plates to near 50% confluency before treatment with 20 µg/mL of WSNTs . After 24 hours of treatment, the cells were washed with Cacodylate buffer and fixed with a 3% glutaraldehyde and 1% paraformaldehyde fixative solution. The cells were then gently scraped off and post-fixed with 2% Osmium Tetroxide, 0.7% Potassium ferrocyanide in Cacodylate buffer. After post-fix, the cells were washed with additional Cacodylate buffer and distilled water before en bloc staining with 1% aqueous uranyl acetate. Finally, the cells were dehydrated with an acetone gradient series, infiltrated into an Epon-Araldite resin and transferred to embedding molds for polymerization. Semi-

thick 400 nm sections of the samples were cut with a UC6 Ultramicrotome (Leica Microsystems) using a diamond knife (Diatome, Ltd.). The sections were placed on formvar-coated, copper-rhodium 2 mm slot grids and stained with 3% uranyl acetate and lead citrate. The grids were placed in a dual-axis tomography holder (Model 2040, EA Fischione Inc.) and imaged with a Tecnai TF30ST transmission electron microscope (ThermoFisher Scientific) at 300 keV. Images were recorded digitally with a US1000 CCD camera (Gatan, Inc.). Tomographic tilt-series data were acquired as described in the literature[34]. Briefly, the grids were tilted ±64° and images were acquired in 1° increments. The grid was then rotated 90° and a similar series was taken about the orthogonal axis. Tomographic data and projection images were processed and analyzed using the IMOD software[35,36] package on a MacPro Computer (Apple, Inc).

Statistical analysis

All biocompatibility statistical tests were performed in GraphPad Prism 7.00. The variation in cellular viability between different conditions was tested using two-way ANOVA for WST-1 assays (statistical significance threshold set at p-value < 0.05). In the nanoparticle concentration and time series studies, statistical tests were performed against untreated cells. In the disk overlay study, statistical tests were performed against cells exposed to PLA.

Preparation of PLA and PLA-WSNT ingots

Vacuum dried PLA pellets and PLA-WSNT films were compression-molded into an ~8 g cylindrical “ingot”, which is the starting material for the flow-induced crystallization

experiments described below. Heat and pressure are applied to the mold using a Carver hot-press. The two platens of the hot-press are maintained at 200°C and the temperature of the mold is continuously recorded used a thermocouple. The mold is also connected to vacuum line (< 300 mTorr) to remove any moisture that can induce degradation of PLA near the melt. The mold is heated up to 180°C and held at that temperature for 10 mins under a pressure of ~2 tonnes. The mold is then removed from the hot-press and rapidly quenched to ~60°C in dry ice. The ingot is subsequently extracted from the mold and stored under vacuum at 40°C.

In situ flow induced crystallization (FIC)

A custom-built apparatus[37] was used to probe the impact of WSNTs on the microstructure of PLA during flow. PLA and PLA-WSNT ingots are first placed in the instrument's reservoir, which is maintained at 200°C, above the melting temperature of PLA (T_m ~ 170°C). A narrow rectangular capillary termed the "shear cell" (6.35 cm x 6.35 mm x 500 μm) is located downstream of the reservoir; the shear cell is filled with material using a pressure-driven piston at the lowest possible wall shear stress to avoid pre-orientation of material. A pressure transducer located near the inlet of the shear cell enables calculation of the wall shear stress (σ_w; see equation S1). The shear cell is then maintained at 200°C for an additional 5 minutes to erase thermal history. The shear cell is cooled to the prescribed shear temperature (Ts: 125 – 140°C) at ~8.5°C/min and a shear pulse (t_s: 10–40 s; σ_w: 0.11–0.23 MPa) is applied using the pressurized piston. A pair of quartz windows at the outlet of the shear cell permit a path for plane polarized light (652 nm He-Ne laser) to interact with the material both during and post flow. A pair of detectors continuously record the transmitted intensity, which is used to

calculate the “retardance”, an optical property analogous to birefringence that provides a measure of oriented crystallization.

Gel permeation chromatography (GPC)

The molecular weight (M_w) of PLA and PLA-WSNT was measured using a Wyatt DAWN EOS Multi-Angle Light Scattering System (MALLS, λ = 690 nm) in conjunction with a Waters 410 differential refractometer (λ = 930 nm). The system uses degassed tetrahydrofuran (THF) as the mobile phase at a temperature and flow rate of 35°C and 0.9 mL/min respectively. The samples are first dissolved in THF at a concentration of 5 mg/mL and the resulting solution is filtered through a 0.45 μm pore poly(tetrafluoroethylene) (PTFE) membrane. The filtered solution is injected into the system and separation is achieved (longest molecules elute first) using four Agilent PLgel columns (pore sizes: 10^3, 10^4, 10^5, and 10^6 Å) connected in series. Elution is complete in 50 minutes and data (light scattering and refractive index) are acquired at a resolution of 5 Hz. Data analysis is performed in the Wyatt Astra software (version 5.3.4) using the Zimm fitting formula with dn/dc = 0.042 ml/g for poly L-lactide. Cognizant that thermal degradation of PLA can alter our interpretation of the retardance data, we checked the M_w of the starting resin, the compression-molded ingots and the extruded material at the end of each flow-induced crystallization experiment. We found that our samples have good thermal stability during compression-molding and for the duration of the shear experiments (~4 hours) as little to no change in M_w was observed (**Figure S2–S3**).

Sectioning and Microscopy

Sheared PLA and PLA-WSNT samples are first embedded in OCT (optimal cutting temperature) media to facilitate sectioning. The embedded samples are microtomed using Sakura Finentek's Tissue Tek-Cryo$_3$ into ~50 µm thick sections both normal to the flow-direction ($\nabla x\mathbf{v}$–$\nabla\mathbf{v}$) and the vorticity-direction ($\mathbf{v}$–$\nabla\mathbf{v}$). The samples are sectioned at -35°C to minimize the impact of the stainless-steel knife (Sakura Accu-Edge 4685 blades) on the morphology of the samples. The microtomed sections are subsequently imaged at 4x through crossed-linear polarizers in a Zeiss Universal Microscope equipped with a Canon EOS DS30 camera.

X-ray scattering

The morphology of PLA and PLA-WSNT samples subjected to shear flow was probed using X-rays at beamline 5-ID-D of the Advanced Photon Source (APS), Argonne National Labs. The incident X-ray beam with spot-size 250 µm x 250 µm was aligned parallel to the gradient ($\nabla\mathbf{v}$), flow ($\mathbf{v}$) and vorticity-direction ($\nabla x\mathbf{v}$) to obtain a 3D-view of the morphology in each sample. Wide Angle X-ray Scattering (WAXS) patterns were acquired on a Rayonix CCD detector that was located 200.83 mm from the sample. Diffraction patterns were acquired with an exposure time of 0.5s using X-rays of wavelength 0.7293Å. Drift in the background scatter was monitored by periodically acquiring "air" scattering patterns (no sample in the path of the beam) at 0.5s exposure as well. The air frames were averaged to obtain a single background image that was subtracted from the sample images to isolate the scattering from

PLA/PLA-WSNT alone. The subtracted patterns were then normalized by the total number of counts to account for variations in thickness.

6.3 Results

In vitro biocompatibility assays were performed on bare WSNTs and PLA-WSNT nanocomposites with appropriate controls in two cell lines – Human Umbilical Vein Endothelial Cells (HUVECs) and Human Aortic Smooth Muscle Cells (HASMCs), which represent the major cell types that are likely to surround the deployed scaffold in the artery[38]. We performed a series of cell culture experiments to assess the biocompatibility of WSNTs and the PLA-WSNT nanocomposites. To evaluate cell tolerance of bare WSNT, we used four methods: phase contrast microscopy[39], Live/Dead staining[40], WST-1 (water soluble tetrazolium with iodo, nitro and disulfo functional groups)[27] assay, and transmission electron microscopy (TEM)[26]. The cytotoxicity of PLA-WSNT nanocomposite films in direct contact with HUVECs and HASMCs for 24 hours was assessed using phase contrast microscopy, Live/Dead staining and WST-1 assays. As an additional test, media that had been incubated with PLA-WSNT nanocomposites for 24 hrs was applied to cells and their metabolic activity was monitored at 24, 48, and 72 hrs. Based on the encouraging results of these biocompatibility assays, we examined the effect of WSNT on morphology development of PLA during processing using short term shear experiments[37,41].

<u>HUVEC and HASMC tolerate bare WSNT *in vitro*</u>

Consistent with the prior literature on the biocompatibility of WSNTs with rat salivary cells[26], human bronchial epithelial cells[27], human hepatocytes[27] and mouse macrophages[27], phase contrast microscopy of HUVECs and HASMCs indicates that both cell lines retain their morphology when exposed to concentrations ranging from 5 to 100 µg/mL of WSNTs (see **Figure 1A** for a schematic of the experiment, 1B-C for microscopy images of 20 µg/mL and **Figure S6A** and **S7A** for other concentrations). Micrographs of cells subjected to Live/Dead staining confirm that cells which retain their expected morphology are indeed alive (**Figure S8**). In negative controls that received ZnO nanoparticle treatment, which is known to be cytotoxic[42,43], both cell lines experience cell death at concentrations above 20 µg/mL of ZnO nanoparticles, (**Figure 1B-C** and **Figure S6B** and **S7B**). Wells containing 20 µg/mL of ZnO were dominated by detached and possibly apoptosed cells (black arrows, **Figure 1B-C**); it was difficult to find any cells with normal morphology in wells exposed to 50 and 100 µg/mL of ZnO (**Figure 1B-C** bottom; and **Figure S6B** and **S7B**).

WST-1 assays of the average metabolic activity performed after a 24-hour treatment with bare WSNTs show that WSNT treated cells, both HUVECs and HASMCs, retain at least 70% of their metabolic activity relative to media controls (WS_2, **Figure 1D-E**). As there is no evidence of cell death at 100 µg/mL of WSNTs (**Figure S6A** and **S7A**), it appears that WSNTs may be influencing the measured metabolic activity by reducing the metabolic and/or cell proliferation rates. In agreement with the micrographs (ZnO, **Figure 1B-C**, *bottom*; and **Figure S6B** and **S7B**), WST-1 measurements show a sharp drop in metabolic activity for ZnO at 20µg/mL, and negligible activity remains for cells exposed to 50 and 100

μg/mL ZnO (**Figure 1D-E**). The results were confirmed by replicate experiments at 20 μg/mL for three time points (24, 48, and 72 hrs) that compared the vehicle control with WSNT and ZnO (**Figure S9**).

HUVEC and HASMC tolerate direct contact with PLA-WSNT nanocomposites *in vitro*

Phase contrast micrographs of cells in contact with PLA-WSNT (0.1 wt%) films for 24 hours show that both HUVECs and HASMCs have similar cell morphology and cell density as their respective media controls underneath the disk (**Figure 2B-C**, *left*). Surprisingly, all three polymeric samples tested—including the PLA-ZnO negative control—had cell density underneath and next to the nanocomposite similar to that of the control untreated cells (**Figure S10**). Since a cytotoxic material would induce cell death at the cell-material interface, constant cell density at the PLA disk interface suggests that the PLA matrix can sequester even a toxic nanoparticle under these testing conditions.

Dead (red) and live (green) channel composite micrographs (right side of each pair of images in **Fig 2B, C**) indicate that all disk treatment conditions have comparable cell viability to the vehicle controls. Cell density appears to be consistent between conditions with only a few dead (red) cells detected. A separate set of samples placed in contact with a PLA or nanocomposite disk overlay was used for WST-1 assays of cell metabolic activity. In both HUVEC and HASMC samples, no detectable difference in metabolic activity was found between PLA compared to PLA-WSNT and PLA-ZnO (**Figure 2D-E**). A minor but statistically significant drop in metabolic activity was detected in HASMC between the

control and PLA. Additionally, these results were confirmed by experiments in which the medium was incubated with PLA nanocomposites for 24 hrs and then applied to cells in culture, which were analyzed at three incubation time points (24, 48 and 72 hrs; **Fig S9**).

TEM of cells exposed to bare WSNT suggest endocytosis and association with cytosolic vesicles

Transmission electron micrographs (TEM) of HUVECs and HASMCs treated with bare WSNTs show black features that are aggregates of WSNTs, intracellular gray regions where the protein or nucleic acid concentration is typical of cytoplasm or nuclear contents, intracellular pale gray to white regions containing vesicles, and pale gray outside the cells where there is embedding resin (**Figure 3A-B**). Of the 91 WSNT aggregates in HUVEC cells and 65 WSNT aggregates in HASMC cells identified in the TEM images, none were observed in the nucleus. The majority of intracellular WSNTs – ~84% (76/91) of the aggregates in HUVECs and ~85% (55/65) of the aggregates in HASMCs – are visibly associated with vesicles. In many of the TEM micrographs, the WSNT aggregates appear to be incompletely surrounded by vesicles (**Figure 3A-IV**). We attribute this observation to the inherent 3D pleomorphic structure of cellular compartments and the inability of TEM to accurately discern the borders of the vesicles in the 2D projection images. To confirm that WSNTs are indeed surrounded entirely by vesicles, tomographic reconstructions were prepared at selected nanoparticle locations to visualize the sample in 3D. Although tomography was limited to a smaller number of WSNT aggregates, the 3D images provided a confident delineation of the vesicle border relative to the nanoparticles within it. In these

reconstructions, the nanoparticle aggregates appear to be fully enclosed by the associated vesicles in both HUVECs and HASMCs (Movie S1 and S2).

Probing the impact on WSNTs on the microstructure of PLLA during flow

We probe changes in microstructure during and after flow with millisecond time resolution by measuring the optical retardance. The average retardance is measured using the variations of light intensity transmitted through crossed and parallel polarizers, with the polarizer and analyzer oriented at ±45° relative to the flow direction and with the laser pointed along the velocity gradient direction[37]. The average retardance is related to the integral of the birefringence over the thickness of the channel[44,45]. Here, there are three contributions to the birefringence: the melt flow birefringence, the birefringence due to formation of oriented PLA crystals, and the birefringence due to oriented nanotubes.

First, consider the retardance during a 10s shear pulse of σ_w = 0.19 MPa as a function of shear temperature (Ts). The application of a flow field distorts the initially isotropic melt, which manifests as an increase in retardance (δ) for the duration of the shear pulse (inset of **Figure 4A**). For a fixed wall shear stress, decreasing Ts mildly increases the retardance during flow. In the absence of nanoparticles (**Figure 4A**, *left*), this increase is indicative of the increase in the melt normal stresses relative to the (fixed) shear stress. For PLA-WSNT, the retardance during the shear pulse accords with the PLA melt retardance (**Fig 4A**, compare *center* to *right* insets), indicating that the contribution of the nanotubes is minimal. A new feature appears when the WSNT concentration is increased to 0.1wt%: near the end of the

shear pulse, the retardance rises steeply for Ts = 133, 130, and 127°C. Furthermore, the behavior after cessation of flow (t = 10s, **Figure 4A**, *right*) is incomplete and the surviving retardance after cessation of shear increases when Ts is lower than 135°C. This occurs despite the fact that the wall shear rate and the total shear strain decrease as Ts decreases (**Figure S12**, *left*).

Next, consider the retardance after cessation of flow as a function of shear temperature (Ts) for matched growth conditions. After the melt relaxes (abrupt decrease in retardance upon cessation of shear), any subsequent increase in retardance is indicative of oriented crystallization templated by oriented precursors and/or oriented nanotubes. In every experiment, the retardance due to crystallization goes over orders (orders are marked by gray horizontal dashed lines, labelled in **Figure 4C**, *left*). Open circles mark the time each order of retardance is reached; the last full order observed is the last data point shown (limited by loss of transmitted intensity due to crystallization). In PLA alone, the increase in retardance due to oriented crystallization reaches the first order for Ts $\leq$ 137°C, indicating oriented precursors were created within 10s during flow (there are no nanotubes). Even though the behavior during the shear pulse is similar for PLA and PLA-WSNT (insets **Figure 4A**), the subsequent increase in retardance due to crystallization is significantly affected by WSNT: the increase in retardance after flow begins earlier and reaches higher values for the nanocomposites than for neat PLA, suggesting that WSNTs promote oriented crystallization in PLA (see t > 10s, **Figure 4A**). The effect of shear temperature shows that deeper

subcooling during the shear pulse results in stronger oriented growth (the rise in retardance begins earlier and the retardance reaches higher values).

To examine the effect of shearing time (t_s: 10–40s, **Figure 4B**) and shear stress (σ_w: 0.11–0.23 MPa, **Figure 4C**), we chose a fixed shearing temperature (T_s = 130°C) that gives a significant retardance after cessation of flow and low enough viscosity to permit flow for 40s even at the highest wall shear stress (σ_w = 0.23 MPa). Increasing the shearing time (t_s) increases the strength of oriented crystallization (**Figure 4B**). This increase interacts with nanotubes: during sustained flow the shear pulse induces an "upturn" (labeled in **Figure 4B**, *mid*) in the retardance at $t \geq 30$s, which increases significantly with the addition of WSNTs. Further, the upturn during shear correlates with significant features after cessation of flow: the retardance does not relax completely and the residual retardance is roughly proportional to the height of the upturn (see **Figure 4B** and **Figure 5**, *left*). In the literature, an upturn in the retardance is associated with a population of shish that survives melt relaxation and promotes growth of kebabs[45–48], as evidenced by an earlier and strong increase in retardance after cessation of flow (**Figure 4B**). For PLA without nanotubes, increasing t_s from 10s to 40s shortens the time to reach the first order in retardance ($\delta = \pi$) by 50s and increases the growth rate of retardance. With the addition of nanotubes, the effect of t_s lessens: increasing t_s from 10s to 40s reduces the time to reach first order only by ~20s for PLA-WSNT 0.05 wt% and <10s for PLA-WSNT 0.1 wt%. Intriguingly, the transient retardance after flow becomes insensitive to t_s at the highest WSNT concentration (0.1 wt%): the time required to observe the first four orders of retardance ($\delta = \pi$, 2π, 3π and 4π) is nearly independent of t_s

(**Figure 4B**, *right*). The effect of the wall shear stress (σ_w) reveals that WSNTs decrease the critical stress required for nucleation of oriented precursors; the stress required to induce an upturn decreases from $\sigma_w \geq 0.17$ MPa for PLA to $\sigma_w \geq 0.13$ MPa for PLA-WSNT 0.1wt% (**Figure 4C**). The interaction of shear stress with nanotubes in producing oriented precursors is also evident in the magnitude of the upturn. For example, at $\sigma_w = 0.19$MPa, the height of the upturn increases 5-fold with addition of 0.1wt% WSNT (**Figure 5A**, *right*)

PLA and PLA-WSNT samples were extracted from the instrument and microtomed at -35°C both normal to the flow (sections in the vorticity–velocity gradient plane, denoted by $\nabla x\mathbf{v}$–$\nabla\mathbf{v}$) and normal to the vorticity (sections in the flow–velocity gradient plane, denoted by $\mathbf{v}$–$\nabla\mathbf{v}$) for *ex situ* polarized light microscopy; therefore, pairs of images are shown for each composition and shearing time that are reminiscent of the classic skin–core morphology observed in injection molding of semicrystalline polymers[49]. The PLA sections are crystalline near the walls and predominantly amorphous towards the core (**Figure 6A**). The lack of significant crystallization in the core indicates the absence of quiescent nucleation events prior to removal of the shear cell from the heater block (at 1000s for all samples); the material in the shear cell then cools into glass within the next 300s. In contrast, the PLA-WSNT sections have a completely crystalline, spherulitic core, indicating that WSNTs act as heterogenous nucleation sites even in the absence of flow (**Figure 6B–C**). Consistent with the *in situ* retardance data (**Figure 4B**), we observe an increase in the retardation of the fully solidified samples with increasing t_s (see color scale[50] and corresponding retardation below **Figure 6**). Intriguingly, the increase in retardation with t_s is more apparent in the plane

normal to the flow direction ($\nabla x\mathbf{v}$–$\nabla\mathbf{v}$, **Figure 6A-C**, *left*) than in the plane containing the flow direction ($\mathbf{v}$–$\nabla\mathbf{v}$, **Figure 6A-C**, *right*). For example, in PLA-WSNT 0.05 wt%, increasing t_s results in a thicker birefringent skin that is readily evident in the $\nabla x\mathbf{v}$–$\nabla\mathbf{v}$ plane (orange-red Michel Levy color, **Figure 6B**, *left*), but there is little to no indication of an oriented skin in the $\mathbf{v}$–$\nabla\mathbf{v}$ plane (**Figure 6B**, *right*). PLA-WSNT 0.1 wt% is the only exception to this trend; in addition to an oriented skin in the $\nabla x\mathbf{v}$–$\nabla\mathbf{v}$ plane (**Figure 6C**, *left*), a weak but distinct skin in the $\mathbf{v}$–$\nabla\mathbf{v}$ plane is observed as well (for t_s: 30–40s, **Figure 6C**, *right*).

6.4 Discussion

WSNT and PLA-WSNT nanocomposite biocompatibility

WS_2 nanoparticles possess beneficial physical and mechanical properties[51] and have recently been studied for various biomedical applications such as friction-reducing agents in nickel-titanium alloys[52] and for reinforcing orthopedic implants[31]. The first step in the use of these nanomaterials for any medical application is to assess their biocompatibility in relevant cells. In a study that explored the use of WSNTs to reinforce salivary gland scaffolds, the nanotubes were introduced to rat submandibular gland-derived A5 cells and found to have no effect on the rate of cell proliferation for concentrations up to 35.2 μg/mL[26]. Similarly, another study investigated WSNT biocompatibility in three different cell types (NL-20 human bronchial epithelial cells, human liver-derived HepG2 cells and mouse Raw264 macrophages) and found high cell survival rate with WSNT concentrations of up to 100

μg/mL[27]. In this study, we are interested in interrogating the potential cytotoxicity of WSNTs for use in a PLA nanocomposite scaffold that can be inserted into occluded arteries for structural support. We found that WSNTs are well tolerated in two relevant human primary cell lines, HUVECs and HAMSCs, for concentrations up to 100 μg/mL after 24 hours of exposure. Both cell lines show a modest, dose-dependent reduction of metabolic activity relative to controls: exposure to 20 μg/mL WSNT (for exposure up to 72 hrs, **Figure S9**) had no measurable effect, but exposure to 50 μg/mL and 100 μg/mL shows a measurable decrease of ~11-24% and up to ~32% respectively (**Figure 1D-E**). We did not observe an abnormal number of floating dead cells at the higher concentrations (100 μg/mL of WSNT), suggesting that WSNTs have a relatively mild impact on cellular metabolic and proliferation rates. Our choice of cytotoxic control (ZnO nanoparticles) was based on prior literature showing strong cytotoxicity of ZnO towards human astrocyte-like U87 cells[43] and human cardiac microvascular endothelial cells[42]. In the present study on HUVECs and HASMCs, we establish a similar dosage-dependent cytotoxicity profile (24 hr, **Figure 1D-E**; up to 72 hr, **Figure S9**). Transmission electron micrographs of cells treated with 20 μg/mL of WSNTs indicate that the nanotubes form aggregates that are endocytosed and enclosed within cytosolic vesicles (**Figure 3**); within the limitations of the image contrast and resolution, no discernable effect on the nucleus, mitochondria, or endoplasmic reticulum was observed.

We also tested the biocompatibility of PLA-WSNT (0.1 wt%) nanocomposite films and found no indication of cytotoxicity with phase contrast microscopy, live/dead staining, and WST-1 metabolic assays. Comparison of the PLA control to PLA-WSNT in HUVECs and

HAMSCs after 24 hours of exposure showed: a) consistent cell density in the central open area and at the cell-film interface (**Figure S10**); b) consistent cell density under the film (**Figure 2**); c) indistinguishable, high viability of cells in the area that was underneath the film (**Figure 2**); and d) comparable metabolic levels as measured by WST-1 assays. Relative to cells that did not have a film placed on them, HASMCs (but not HUVECs) show a small, statistically significant decrease in metabolic activity, independent of WSNT content (**Fig 2E**). Likely causes include reduced nutrient transport due to disk-overlay and/or disruption of cells due to insertion and removal of disks. Interestingly, the PLA-ZnO (0.1wt%) nanocomposite has favorable biocompatibility with little difference from the PLA control (**Figure 2**). This may be due to negligible hydrolysis of PLA during the treatment period (24 hours), such that the ZnO nanoparticles remain trapped in the PLA matrix and consequently, have no detectable impact on cellular metabolic activity.

In situ oriented crystallization in PLA-WSNT relative to PLA

The literature on flow-induced crystallization is scarce for PLA, particularly for PLA-nanocomposites. The majority of studies on the effects of shear on PLA morphology, particularly those that observe structure development during shear, impose shear flow using a parallel-plate geometry (Linkam CSS-450, shear rate: $\dot{\gamma}$ = 1–100s^{-1})[53–57]. Prior studies of flow-induced crystallization of PLA include some that examine polymers similar to the present study (references 53-56 use PLA from same supplier, with similar <2% D-content and M_w). Interestingly, prior works report relatively low densities of oriented structures (row-structures ~50μm apart)[55–57], even for PLA containing 0.1wt% carbon nanotubes (row-

structures ~10μm apart)[56]. In contrast, we observe densely packed oriented structures (<5μm apart, **Figure 6**) despite operating at a much lower shear rate (here $\dot{\gamma}$ <1s^{-1}) and shear duration (t_s: <40s) compared to the studies described above[55–57] ($\dot{\gamma}$: 5 to 30s^{-1}; t_s: 3 to 10 mins). Only at much higher shear rates (10, 25 and 100s^{-1})[53,54] and shear durations (5 to 10 mins)[53] are tightly packed oriented structures (shish-kebabs < 5 μm apart) reported. Although the origin of this stark difference is unclear, a contributing factor might be a difference in molecular weight: the sample handling and experiment conditions in this study preserve the initial molecular weight (**Figure S2–S3**), in contrast to reports of a >20% decrease in the PLA molecular weight[53], or severe degradation[54]. Perhaps retaining high molecular weight species enhances formation of oriented precursors during flow. If the stress observed during prior experiments was known, it might be feasible to test this possibility; however, the stress was not measured in these previous studies[53–57]. In addition to the rapid and highly oriented crystallization of PLA, we also observe a fascinating anisotropy in the oriented skin indicating that the structures are not cylindrulites (i.e., they are not isotropic in the plane orthogonal to the flow direction). The PLA-WSNT nanocomposites exaggerate the unusual behavior: nanotubes further enhance rapid, highly oriented crystallization and the peculiar anisotropy in the oriented skin increases (for increasing WSNT content, compare the bottom row in **Figure 6**, *left column*). The remainder of the Discussion is devoted to these intriguing features evident in our data: (1) why does an increase in WSNT concentration reduce the effect of shear time; (2) what is the molecular basis for the unusual morphology observed in

the PLA/PLA-WSNT micrographs; and (3) how does oriented crystallization vary through the thickness of a sample?

PLA and PLA-WSNT differ in their response to an increase in the shear duration (t_s). For PLA, an increase in t_s increases the slope of the retardance traces after cessation of flow (PLA traces fan out, compare t_s: 10–40s, **Figure 4B**, *left*, 50-200s). The increase in slope is consistent with the expected effect of increasing t_s: the activation of point-like nuclei that form thread-like precursors continues for the duration of the shear pulse and the initial rate of increase of birefringence after cessation of flow is proportional to the total shish length per unit volume in the oriented skin[45]. The addition of WSNTs changes this behavior: the retardance profiles for PLA-WSNT at the end of the shear pulse are more or less unaffected by t_s; the retardance traces are clustered together and have almost identical slopes (see **Figure 4B**, *mid–right*). It appears that an increase in t_s has little bearing on the concentration of thread-like precursors; this behavior is reminiscent of "saturation" of shish in isotactic polypropylene (iPP)[46,48]. This saturation effect in the retardance traces manifests in the *ex situ* micrographs as well: with increasing t_s (**Figure 6A**, *left*), the degree of anisotropy in the layers near the wall clearly increases in PLA alone, but not for t_s >20s in PLA-WSNT 0.05 wt% (**Figure 6B**, *left*) nor for any t_s in PLA-WSNT 0.1 wt% (**Figure 6C**, *left*). The effect of WSNT concentration (cf. PLA-WSNT 0.05 wt% and 0.1 wt% retardance post flow, **Figure 4B**, *mid–right*) suggests that a concentration of 0.1 wt% has enough oriented nanotubes that any shear-induced oriented precursors have negligible effect: the curves are not only parallel, they are almost indistinguishable for all t_s (**Figure 4B**, *right*). If WSNTs are dispersed

uniformly, a 0.1 wt% loading of oriented WSNTs translates to a spacing of ~3.5 μm (see SI for calculation) between adjacent nanotubes, which makes it likely that the nanotubes act as preexisting oriented structures in the PLA melt.

The oriented skin in these samples has a feature we have not found in the shear-induced crystallization of any other semicrystalline polymer: a lack of rotational symmetry about the flow direction. To our knowledge, none of the copious literature on skin-core morphologies reports a deviation from rotational symmetry of the crystallites growing outward from the shish. Shear leads to shish that have the chain axis along the flow direction in isotactic polypropylene and polyethylene, among others. Lamellae nucleate on the shish and grow radially outward. Sections cut in the plane of the gradient and vorticity directions (such that the observer is looking down the flow direction, hence the chain axis in the shish) have an oriented skin that appears dark when viewed between crossed polars; this is usually interpreted to mean that the *a*- and *b*- axes are randomly oriented in the plane perpendicular to the *c*-axis[46]. This morphology is referred to as cylindrulitic. Sections cut in the plane of the flow and gradient directions (observer looking along the vorticity axis) have an oriented skin that appears bright when viewed between crossed polarizers, due to the pronounced difference in polarizability along the *c*-axis compared with the *a*- and *b*-axes of the crystal[46].

Electron microscopy on PLA confirms that formation of shish along the flow direction can be induced by shearing the subcooled melt, albeit for relatively long shearing times and high shear strain[49,53]. However, we were unable to find any mention of an unexpected asymmetry about the flow direction. In the present specimens, the pairs of images in the plane of the

gradient and vorticity (Figure 6A-C, *left*) and in the plane of the flow and gradient (see **Figure6A-C**, *right*) violate the expectation that the oriented skin on the left should show no retardance (here, a full wave retarder is used, and null retardation would appear first order red/purple). Instead, for each material and each shear time, that is the projection that appears bright—consistently brighter than its partner image in the plane of the velocity and gradient (on the right for each pair of images in **Figure 6**). We confirmed that the bright skin seen in (∇x**v**–∇**v**)-plane is due to preferentially-oriented crystallites and not densely packed spherulites by rotating the sample and observing that the skin appears bright when oriented -45° or 45° relative to the crossed polarizers and dark when it is parallel to them (**Figure S19A-C**, *left*). X-ray scattering data acquired along all three projections (**v**–∇x**v**, ∇x**v**–∇**v**, and **v**–∇**v**) confirm that the orientation distribution of crystallites is not symmetric about the flow direction (**Figure 7**). The observed WAXS patterns are a superposition of the scattering from a thin layer of oriented crystallites and a substantially thicker layer of spherulites. In the (**v**–∇x**v**)-plane, all three samples (PLA, PLA-WSNT 0.05%, and PLA-WSNT 0.1%) indicate a population of crystallites with their *c*-axes aligned along the **v**-direction (**Figure 7A**). However, when the same sections are probed in the (∇x**v**–∇**v**)–plane, we observe crystallites with their *c*-axes aligned along the ∇**v**-direction (**Figure 7B**), which may explain the presence of a bright skin viewed in this projection (**Figure 6A-C**, *left* image of each pair). In the (**v**–∇**v**)-plane, there is little evidence of a preferred direction of orientation (**Figure 7C**), consistent with the absence of a distinct skin in the micrographs (**Figure 6A-C**, *right* image

of each pair). Thus, the X-ray data corroborate the micrographs and confirm the absence of cylindrical symmetry in the sheared samples.

The observed *in situ* retardance profiles are depth-averaged measurements as the beam of plane polarized light passes through the entire 500 μm thickness ($\nabla\mathbf{v}$-direction) of the sample. Therefore, we take advantage of the linear shear stress profile in the channel and apply a "depth-sectioning" approach[44] to the series of experiments in which σ_w is varied with T_s and t_s held fixed (**Figure 4C**). The analysis isolates the transient retardance of the material at specific distances from the walls in a sample subjected to σ_w = 0.21 MPa. The depth-sectioned traces for all three materials (PLA with 0, 0.05, and 0.1wt% WSNT) share some common features: the relaxation in retardance after cessation of flow is most pronounced in the layer farthest from the walls (~85 μm, **Figure 8A**); the retardance begins to increase within 5s after cessation of flow at every distance from the walls; the rate of retardance increase is greater the closer the material is to the walls (**Figure 8A to D**, *i.e*. the higher the local stress); and the time at which impingement of neighboring crystals (the inflection point in **Figure 8**) becomes evident is earlier the closer the material is to the wall. All of these features are indicative of greater formation of flow-induced precursors at greater shear stress: oriented precursors prevent complete relaxation of retardance after cessation of shear, increase the number of oriented lamellae that subsequently nucleate, and decrease the distance oriented lamellae grow prior to impingement. At each distance from the wall, adding WSNTs causes the retardance to rise sooner and more steeply than it does for PLA alone. Quantitatively, oriented crystallization develops as rapidly in the PLA-WSNT 0.1wt% layer

farthest from the wall (~85μm, blue line in **Figure 8A**) as it does in the PLA layer closest to the wall (~20μm, black line in **Figure 8D**). This suggests that 0.1wt% WSNTs make up for a substantial stress deficit of ~0.05MPa over a distance of 60μm. The growth velocity is expected to be the same in all of these experiments (that of quiescent pure PLA at 135°C), so the differences in the initial rate of increase of the retardance reveal differences in the number of oriented lamellae that nucleate on shear-induced precursors, and differences in the time at which impingement occurs reveal the distance between precursors. If we consider the layer closest to the wall, the "roll-off" in the slope indicates that impingement occurs at ~100s for PLA and ~70s for PLA-WSNT (~60s post flow for PLA and ~30s post flow for PLA-WSNT, see arrows in **Figure 8D**). Based on the time at which impingement becomes evident and the quiescent lamellar growth velocity of this grade of PLA (~1.8 μm/min at 135°C[55]), we estimate that the separation between shish decreases from ~3.5μm apart in PLA to ~2μm apart in PLA-WSNT (both 0.05 and 0.5%).

6.5 Conclusions

There is an unmet need for a bioresorbable vascular scaffold (BVS) that has thickness comparable to metal stents (~80μm) for the treatment of lesions in smaller and tortuous arteries. The poor radio-opacity of polymers compared to metals is an added complication as surgeons find it challenging to visualize a BVS with X-rays. Towards the goal of a thinner, stronger, and radio-opaque BVS, we evaluate Tungsten Disulfide (WS_2) nanotubes (WSNTs) as a candidate additive to polylactide (PLA) to reinforce PLA and confer radio-

opacity comparable to clinical standards (e.g., platinum markers). Using cell lines (HUVEC and HASMC) that are relevant to vascular tissue, we assess the biocompatibility of bare WSNTs and PLA-WSNT nanocomposites (0.1 wt%) against appropriate controls. Cells treated with the bare WSNTs (up to 100 μg/mL) and PLA-WSNT films retain their morphology and metabolic activity. Transmission electron micrographs of cells exposed to 20 μg/mL WSNTs indicate that the nanotubes are endocytosed and appear to be trapped in cytoplasmic vesicles.

The promising *in vitro* biocompatibility of WSNTs motivated us to explore PLA-WSNT from a materials science perspective. The clinically-approved BVS derives its radial strength from an oriented microstructure that develops under elongational deformation. Therefore, we designed short term shear experiments that probe the impact of WSNTs on oriented crystallization of PLA. *In situ* rheo-optical measurements show that inclusion of 0.1wt% WSNTs reduced both the critical shear duration and the shear stress required for induction of oriented precursors during the short shear pulse. These template oriented crystallization, which can increase strength in semicrystalline polymers, relevant to the goal of enabling thinner vascular scaffolds. *Ex situ* polarized light microscopy and X-ray scattering led to a surprising discovery: in both PLA alone and PLA-WSNT, the shear-induced oriented semicrystalline morphology does not have cylindrical symmetry about the flow direction.

The results presented in this report suggest that WSNTs may be viable reinforcing additives and X-ray contrast agents for biomedical implants. Proposed future biocompatibility studies include *in vivo* testing of the nanocomposite in appropriate animal models. Future *in vitro*

experiments may reveal the degradation kinetics and degradation products of WSNT. The flow-induced crystallization results provide an impetus to prepare PLA-WSNT preforms and to process them into ~80μm thick tubes to study laser cutting and crimping of PLA-WSNT. The connection between microstructure and strength for the PLA-WSNT BVS in the expanded, crimped, and deployed state might open a path toward a future generation of thinner, stronger, and radio-opaque scaffolds.

6.6 Figures and Tables

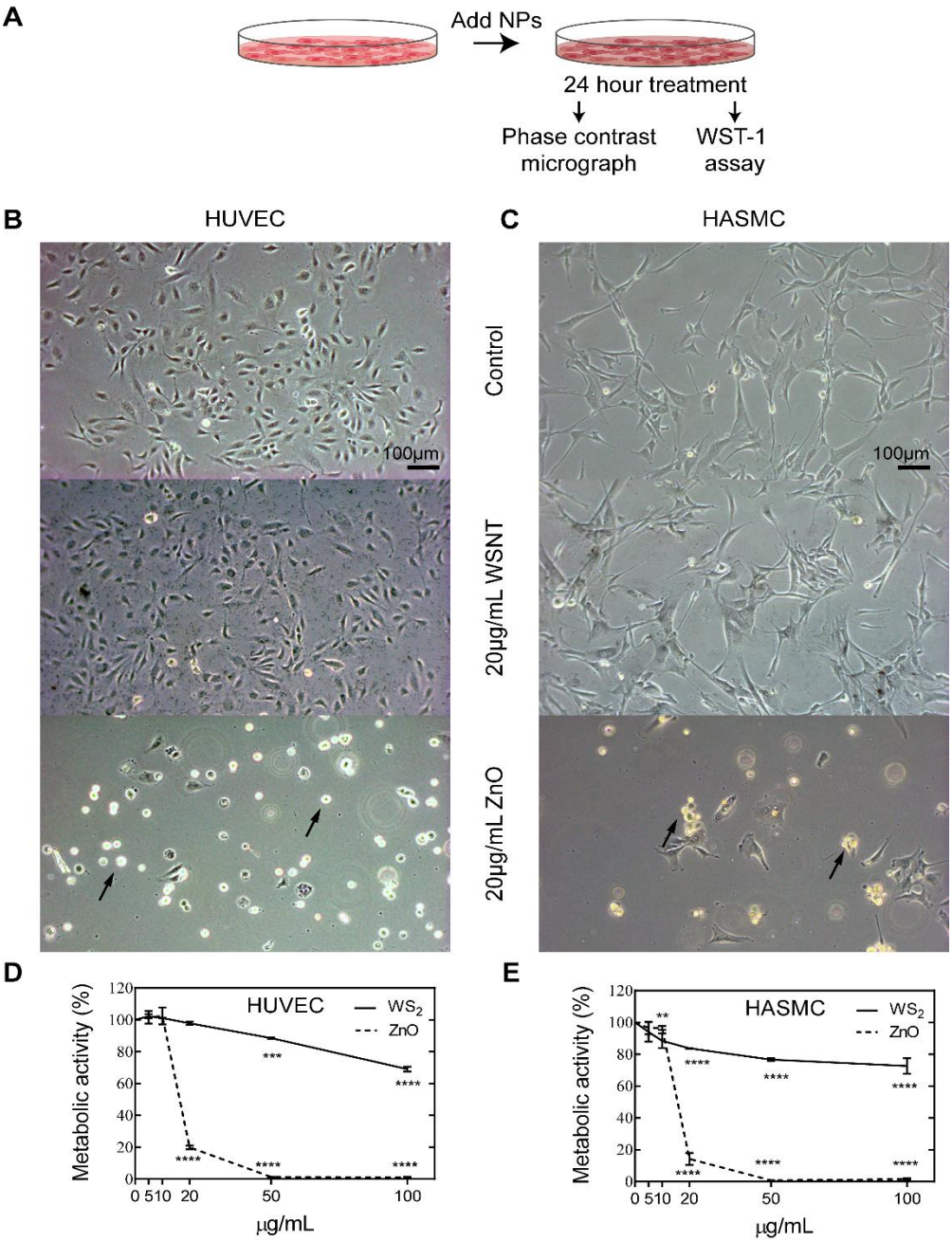

Figure 6.1 Nanoparticle incubation followed by microscopy and metabolic assay
(**A**) Schematic illustrating that cells grown in 6-well plates are treated with varying concentrations of WSNT or ZnO nanoparticles for 24 hours before (**B-C**) phase contrast microscopy and (**D-E**) WST-1 assays. Experiments were performed twice with six replicates

per condition. (**B**) HUVECs and (**C**) HASMCs incubated with 20 µg/mL WSNT retained normal cell morphology while those incubated with 20 µg/mL ZnO underwent cell death (black arrows). (**D, E**) Cells exposed to increasing concentrations of WSNT showed a moderate drop in metabolic activity. ZnO induced strong cytotoxicity at concentrations ≥ 20µg/mL. Mean and standard error of mean plotted. All statistical tests were performed against control samples and stars indicate the magnitude of adjusted p-value (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$).

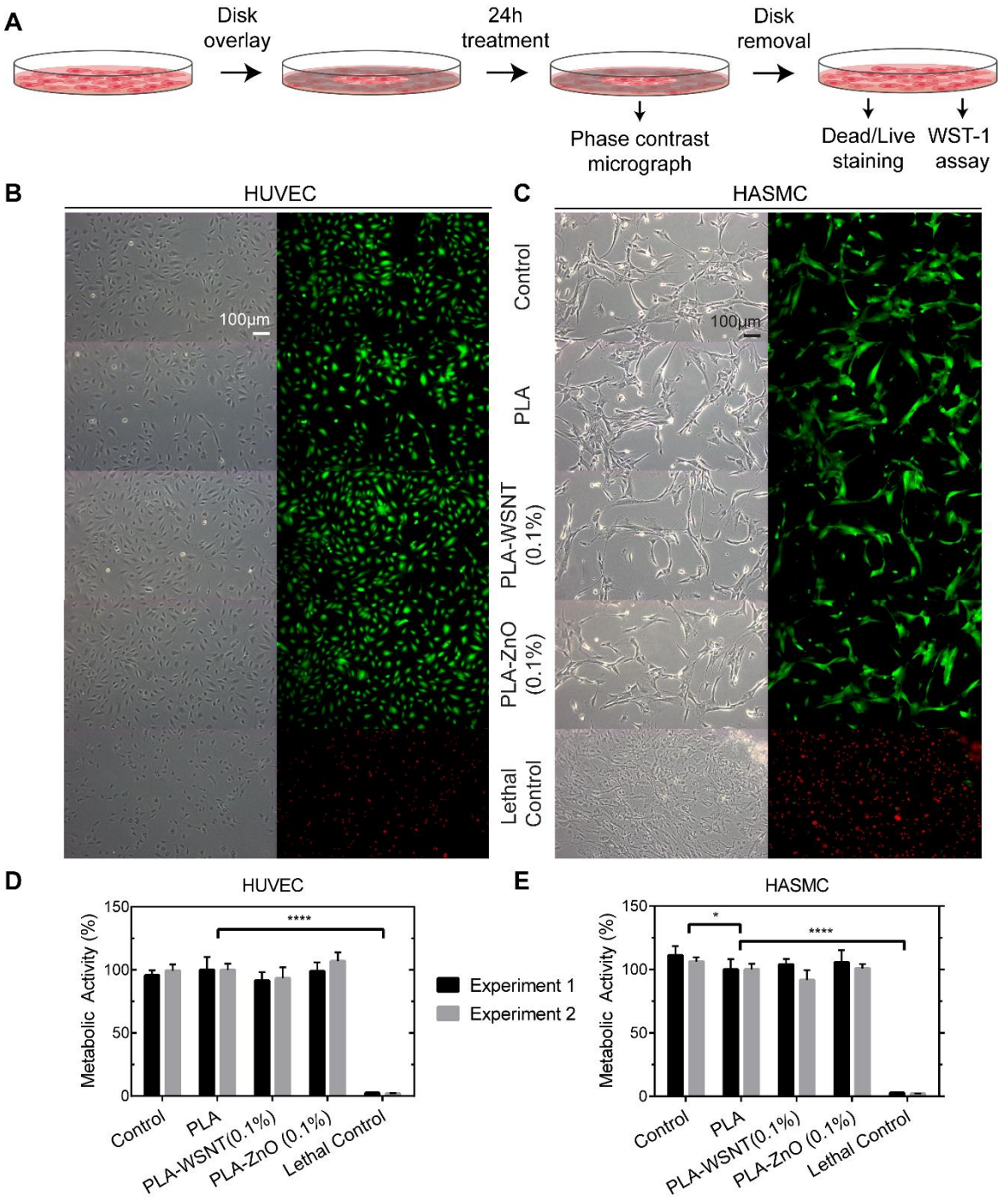

Figure 6.2 PLA disk incubation followed by microscopy and metabolic assay
(**A**) Schematic of HUVECs and HASMCs treated with polymer disks for 24 hours in 12-well plates. (**B-C**, *left*) Phase contrast micrographs and (**B-C**, *right*) merged live (green) dead (red)

stained images of cells underneath the disk show that the treated cells retain their morphology and remain viable. (**D-E**) Cellular metabolic activity levels, inferred from the WST-1 assay, show no statistical difference in viability between cells treated with PLA disks and PLA nanocomposite (WSNT or ZnO) disks. Experiments were performed twice with four replicates per condition. Experiment-specific mean and standard deviation are plotted. All statistical tests were performed against the PLA condition; stars indicate the magnitude of adjusted p-value (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$).

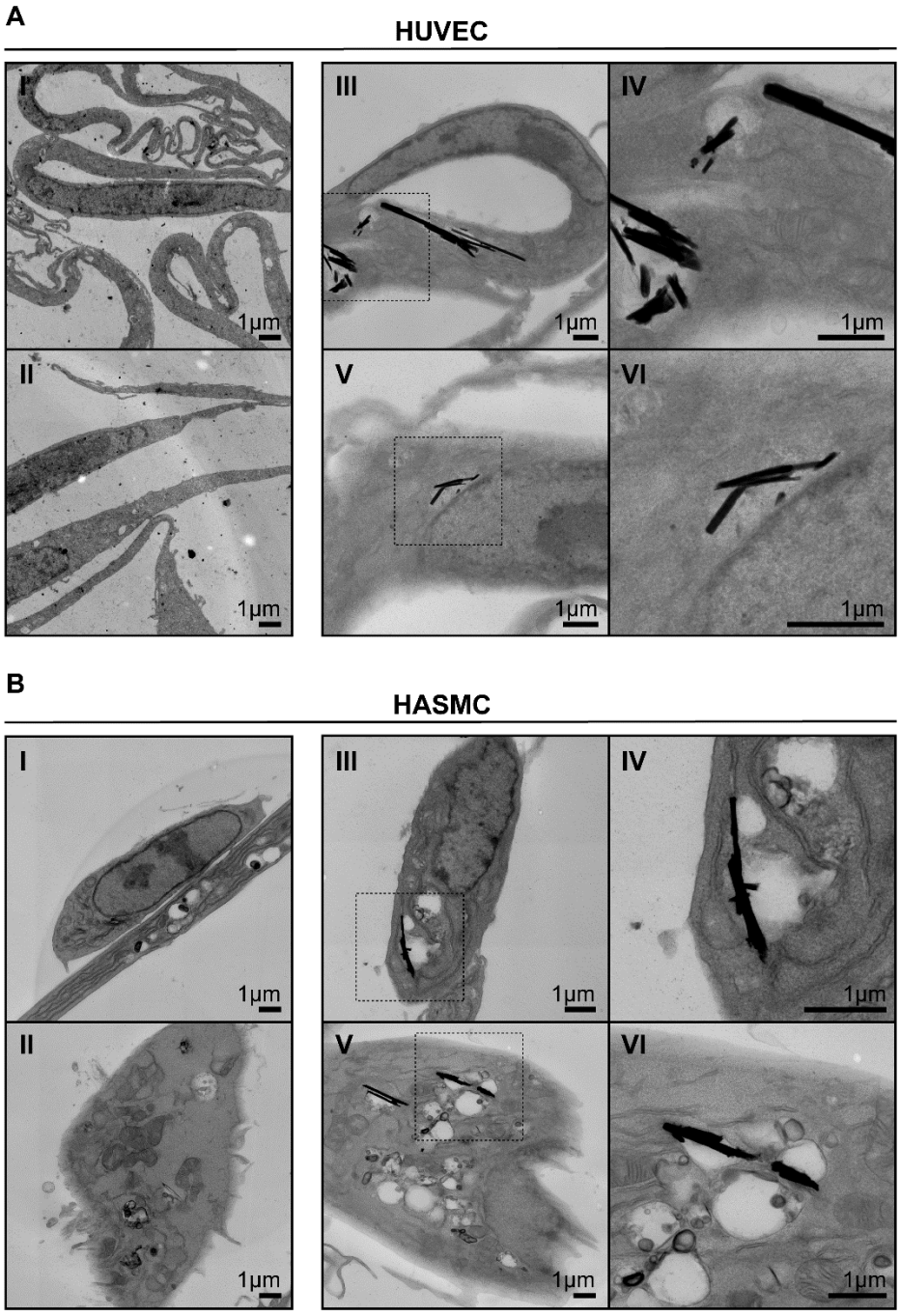

Figure 6.3 TEM of cells with endocytosed WSNT
TEM images of (**A**) HUVECs and (**B**) HASMCs exposed to 20 μg/mL WSNT for 24 hours. In A-B, (I–II) control HUVECs and HASMCs; (III & V) endocytosed WSNTs in HUVECs and HASMCs and (IV & VI) enlarged images of the endocytosed nanotubes in the regions bounded by a dashed box in (III & V).

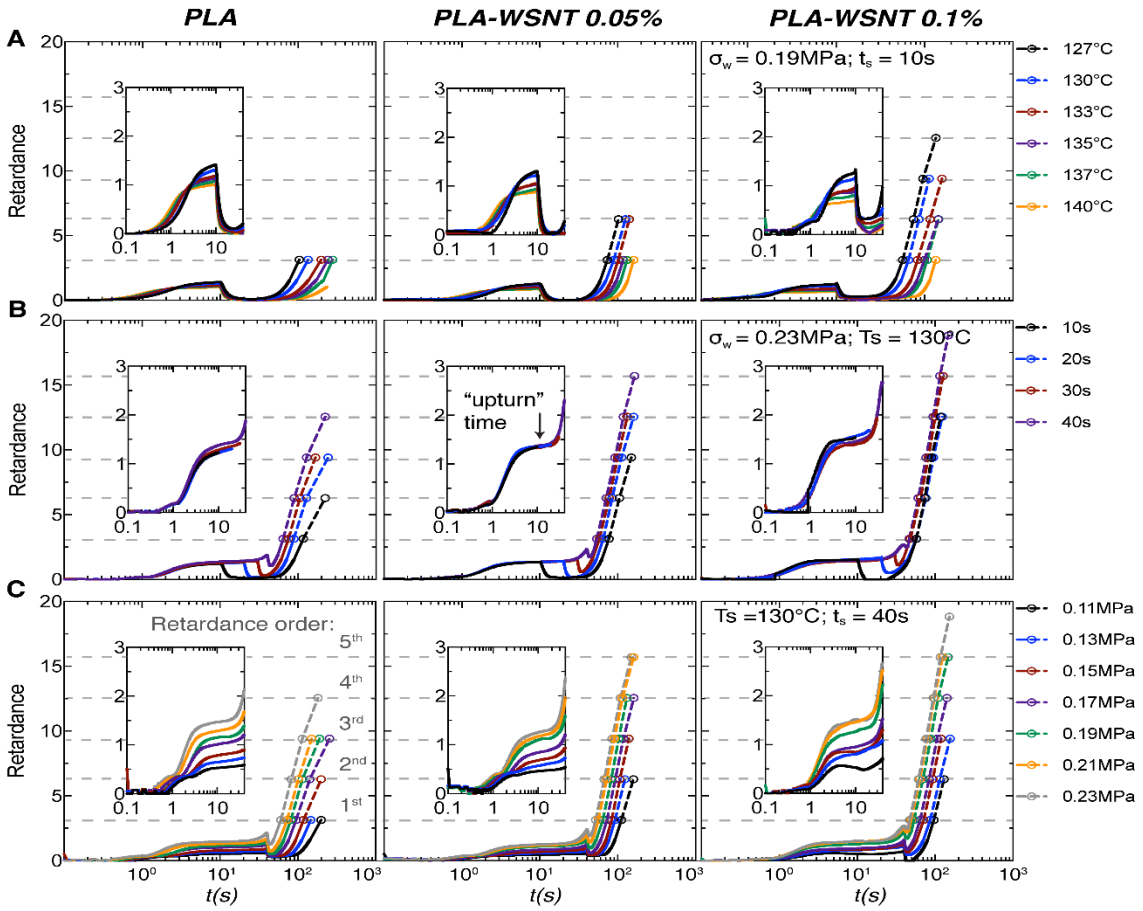

Figure 6.4 *In situ* retardance profiles

In situ retardance profiles during and after cessation of flow for (*left*) PLA, (*mid)* PLA-WSNT (0.05 wt%) and (*right*) PLA-WSNT 0.1 wt% subjected to varying (**A**) shear temperature (Ts: 127 to 140°C), (**B**) shear duration (t_s: 10 to 40s), and (**C**) wall shear stress (σ_w: 0.11 to 0.23 MPa) at matched thermal and flow conditions (Figure S11). In (A), the wall shear rate varies from ~0.15s^{-1} at 125°C to ~0.8s^{-1} at 140°C; in (B), the wall shear rate is ~ 0.35s^{-1} for all cases; and in (C), the wall shear rate varies from ~0.19s^{-1} at 0.11MPa to ~ 0.35s^{-1} at 0.23MPa. The wall shear rate is calculated from the extruded material (Figure S12A) and the Rabinowitsch correction (equations S2–S4) is applied to account for non-Newtonian flow (Figure S12B). The onset of an "upturn" in the retardance during flow is indicated by a black arrow in (B, *mid*). Quantitative characteristics of the upturn for PLA and PLA-WSNT are presented in Figure 5. The dashed horizontal gray lines indicate the order of retardance (δ), which follows integer multiples of π (e.g. 1st order: $\delta = \pi$; 2nd order: $\delta = 2\pi$ and so on). The retardance is calculated from normalized intensity traces presented in Figure

S13–S15. The extruded material from each condition was subject to gel permeation chromatography (GPC) measurements that indicated little to no change in molecular weight for the experiments in A-C (Figure S2-S3).

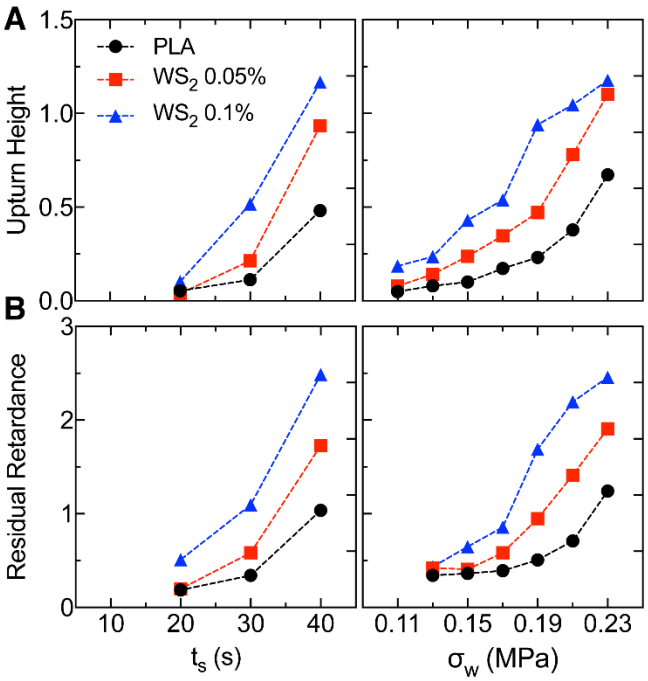

Figure 6.5 Upturn height and residual retardance
In situ retardance traces of PLA and PLA-WSNT were analyzed to compute (**A**) the "upturn height", *i.e.*, retardance at the upturn time (defined in Figure 4B inset) and (**B**) the residual retardance after cessation of flow as a function of (*left*) shear duration (see Figure 4B) and (*right*) wall shear stress (see Figure 4C). We describe the calculation of the upturn height and residual retardance in the SI (Figure S16–17).

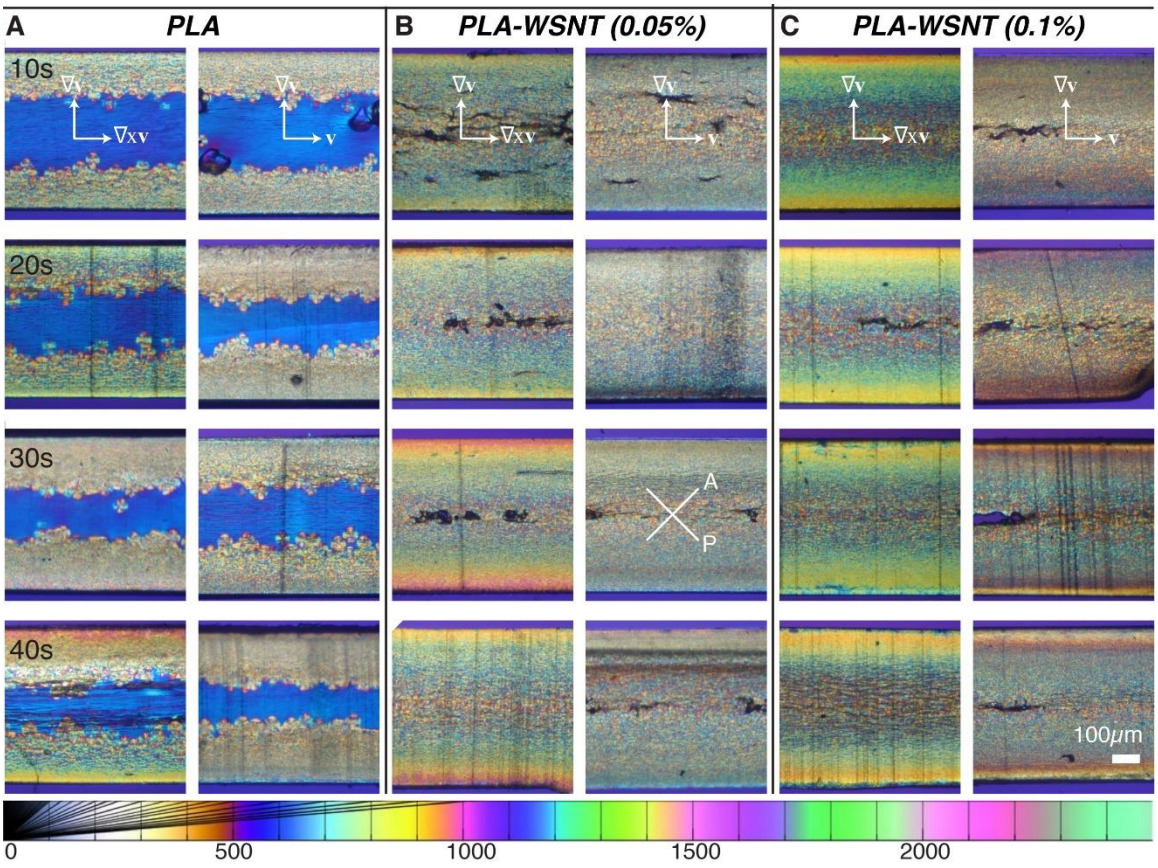

Figure 6.6 Polarized light micrograph
Polarized light micrographs of ~50μm thick sections from samples sheared for ts from 10 to 40s at Ts = 130°C and σw= 0.23MPa for **(A)** PLA, **(B)** PLA-WSNT (0.05 wt%) and **(C)** PLA-WSNT (0.1 wt%). For each composition, pairs of images are displayed: the first column **(A-C, left)** presents micrographs of sections cut normal to the flow direction (vorticity – velocity gradient plane, labelled as ∇xv–∇v) and the second column **(A-C, right)** presents micrographs of sections cut normal the vorticity direction (flow – velocity gradient plane, labelled as v–∇v). Images are acquired through linear crossed polarizers, analyzer **(A)** and polarizer (P) orientations are indicated in **(B)** 30s, right), with a full-wave retardation plate inserted (slow axis vertical, see Figure S18 for details). The Michel-Levy color chart at the bottom (adapted from Ref. [50]) relates colors to retardation in nm.

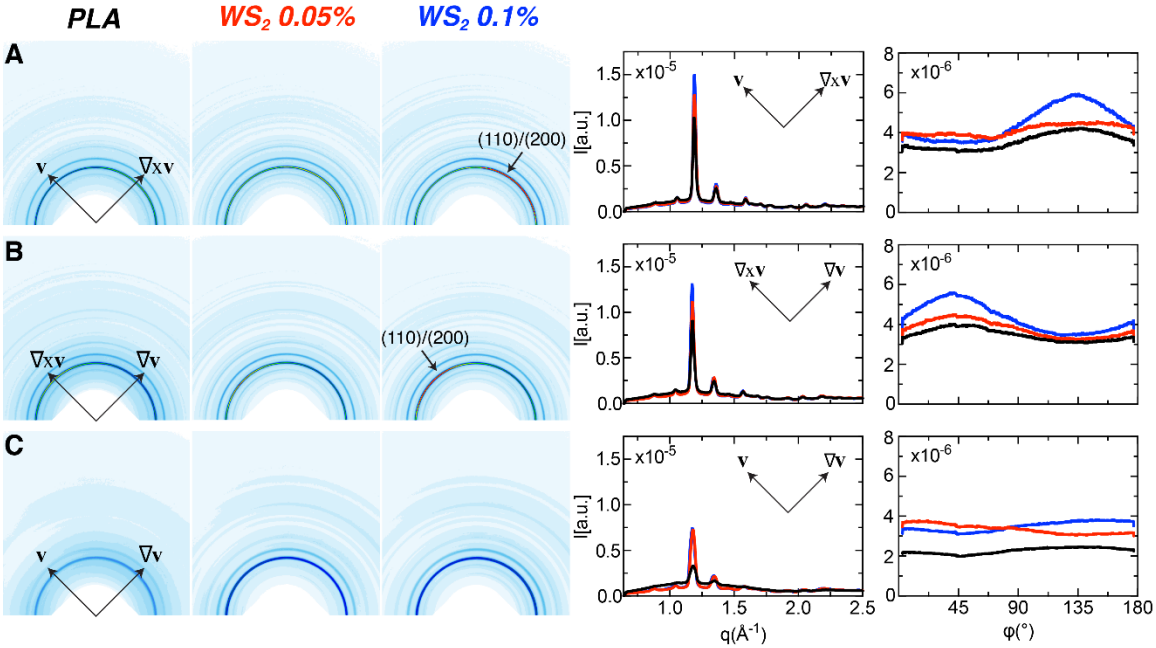

Figure 6.7 Wide Angle X-ray Scattering of PLA and PLA-WSNT

Wide Angle X-ray Scattering (WAXS) data acquired on PLA, PLA-WSNT 0.05wt% and PLA-WSNT 0.1wt% samples sheared at T_s = 130°C, t_s = 40s and σ_w = 0.21MPa. For each sample, the microstructure was probed along all three projections: (**A**) flow – vorticity (**v**–∇x**v**), (**B**) vorticity – velocity gradient (∇x**v**–∇**v**) and (**C**) flow – velocity gradient (**v**–∇**v**). The WAXS data are presented as (A-C, left) 2D patterns and (A-C, right) azimuthally, I(q), and radially averaged, I(φ), intensity plots. Radial averaging for the I(φ) plots is performed near the (110)/(200) diffraction: $1.13 \leq q \leq 1.23$ Å^{-1}. The 1D plots for PLA, PLA-WSNT 0.05wt% and PLA-WSNT 0.1wt% (A-C, right) are presented in black, red and blue respectively. The WAXS patterns are presented using a custom colormap that varies from 0 [white] to 1.75e^{-5} [red] counts. X-ray data acquired at different wall shear stresses (σ_w = 0.11 and 0.15 MPa) for each sample are presented in the SI (see Figure S20 – S22).

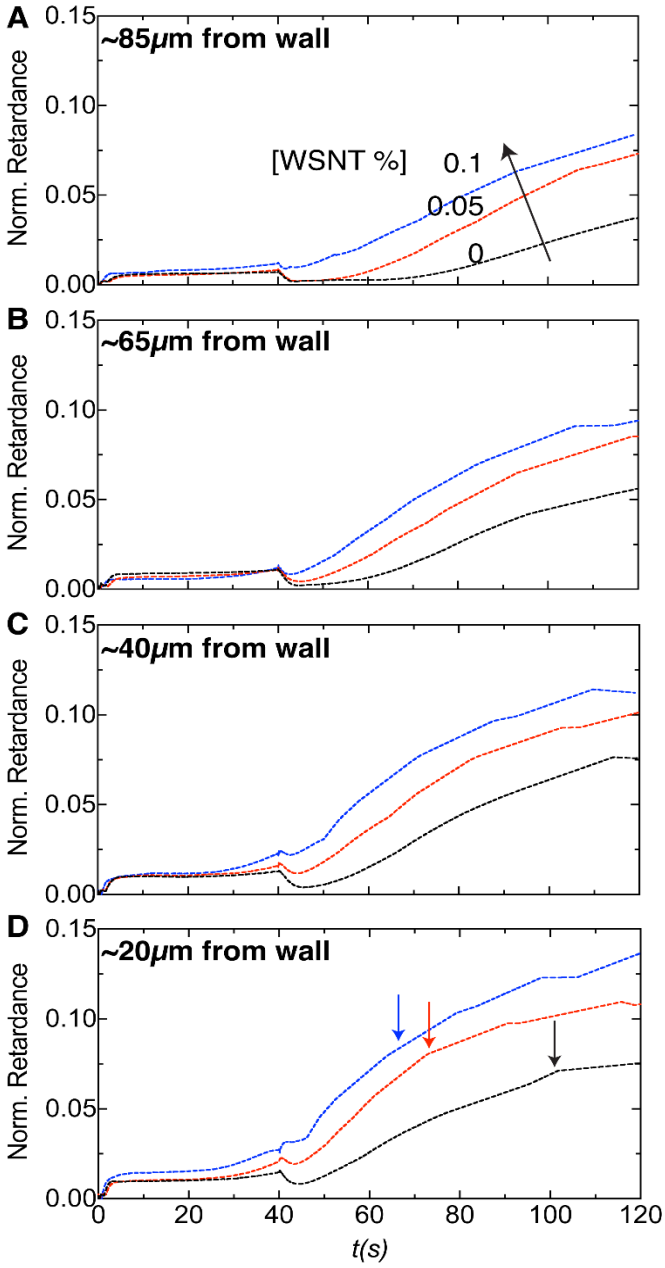

Figure 6.8 Local Retardance

In situ retardance data acquired for PLA (black), PLA-WSNT 0.05wt% (red) and PLA-WSNT 0.1wt% (blue) at varying wall shear stresses (Figure 4C) are "depth-sectioned" to isolate the local retardance at **(A)** 85, **(B)** 65, **(C)** 40, and **(D)** 20 μm from the wall for σw = 0.21MPa. The arrows in (D) mark inflection points. See equations S9–S10 and Table S1 for the depth-sectioning calculations.

6.7 References

1. World Health Organization. *Global Atlas on Cardiovascular Disease Prevention and Control*. (2010).
2. Centers for Disease Control and Prevention. *Health, United States*. (2016).
3. Ding, N., Pacetti, S., Tang, F., Gada, M. & Roorda, W. XIENCE V (TM) Stent Design and Rationale. *Journal of Interventional Cardiology* **22**, S18–S27 (2009).
4. Epstein, A. J., Polsky, D., Yang, F., Yang, L. & Groeneveld, P. W. Coronary revascularization trends in the United States, 2001-2008. *JAMA : the journal of the American Medical Association* **305**, 1769–1776 (2011).
5. Wiebe, J., Nef, H. M. & Hamm, C. W. Current Status of Bioresorbable Scaffolds in the Treatment of Coronary Artery Disease. *Journal of the American College of Cardiology* **64**, 2541–2551 (2014).
6. Ong, D. S. & Jang, I.-K. Causes, assessment, and treatment of stent thrombosis—intravascular imaging insights. *Nature Reviews Cardiology* **12**, 325–336 (2015).
7. Langer, R. & Tirrell, D. A. Designing materials for biology and medicine. *Nature* **428**, 487–492 (2004).
8. Tibbitt, M. W., Rodell, C. B., Burdick, J. A. & Anseth, K. S. Progress in material design for biomedical applications. *Proceedings of the National Academy of Sciences* **112**, 201516247 (2015).
9. Rizik, D. G. & Padaliya, B. B. Early U.S. Experience Following FDA Approval of the ABBOTT Vascular Bioresorbable Vascular Scaffold: Optimal Deployment Technique Using High Resolution Coronary Artery Imaging. *Journal of Interventional Cardiology* (2016). doi:10.1111/joic.12329
10. Kossuth, M. B., Perkins, L. E. L. & Rapoza, R. J. Design Principles of Bioresorbable Polymeric Scaffolds. *Interventional Cardiology Clinics* **5**, 349–355 (2016).
11. Ormiston, J. A. *et al.* First serial assessment at 6 months and 2 years of the second generation of ABSORB everolimus-eluting bioresorbable vascular scaffold a multi-imaging modality study. *Circulation: Cardiovascular Interventions* **5**, 620–632 (2012).
12. Iqbal, J. *et al.* Bioresorbable scaffolds: Rationale, current status, challenges, and future. *European Heart Journal* **35**, 765–776 (2014).
13. Serruys, P. W. *et al.* A Polylactide Bioresorbable Scaffold Eluting Everolimus for Treatment of Coronary Stenosis: 5-Year Follow-Up. *Journal of the American College of Cardiology* **67**, 766–76 (2016).
14. Serruys, P. W. *et al.* A bioresorbable everolimus-eluting scaffold versus a metallic everolimus-eluting stent for ischaemic heart disease caused by de-novo native coronary artery lesions (ABSORB II): An interim 1-year analysis of clinical and procedural secondary outcomes from . *The Lancet* **385**, 43–54 (2015).

15. Hu, Y., Rogunova, M., Topolkaraev, V., Hiltner, A. & Baer, E. Aging of poly(lactide)/poly(ethylene glycol) blends. Part 1. Poly(lactide) with low stereoregularity. *Polymer* **44**, 5701–5710 (2003).
16. Banks, H. T., Hu, S. & Kenz, Z. R. A Brief Review of Elasticity and Viscoelasticity for Solids. **3**, 1–51 (2011).
17. Rathi, S. *et al.* Toughening semicrystalline poly(lactic acid) by morphology alteration. *Polymer* **52**, 4184–4188 (2011).
18. Grijpma, D. W. & Pennings, A. J. (Co) polymers of L-lactide , 1. Synthesis , thermal properties and hydrolytic degradation. *Macromolecular Chemistry and Physics* **195**, 1633–1647 (1994).
19. Huang, M. H., Li, S. & Vert, M. Synthesis and degradation of PLA-PCL-PLA triblock copolymer prepared by successive polymerization of ε-caprolactone and DL-lactide. *Polymer* **45**, 8675–8681 (2004).
20. Li, S. Hydrolytic degradation characteristics of aliphatic polyesters derived from lactic and glycolic acids. *Journal of Biomedical Materials Research* **48**, 342–353 (1999).
21. Xu, J.-Z. *et al.* Graphene Nanosheets and Shear Flow Induced Crystallization in Isotactic Polypropylene Nanocomposites. *Macromolecules* **44**, 2808–2818 (2011).
22. Li, C. Y., Li, L., Cai, W., Kodjie, S. L. & Tenneti, K. K. Nanohybrid shish-kebabs: Periodically functionalized carbon nanotubes. *Advanced Materials* **17**, 1198–1202 (2005).
23. Vardharajula, S. *et al.* Functionalized carbon nanotubes: Biomedical applications. *International Journal of Nanomedicine* **7**, 5361–5374 (2012).
24. Naffakh, M., Marco, C. & Ellis, G. Development of novel melt-processable biopolymer nanocomposites based on poly(l-lactic acid) and WS2 inorganic nanotubes. *CrystEngComm* **16**, 5062 (2014).
25. Adini, A. R., Redlich, M. & Tenne, R. Medical applications of inorganic fullerene-like nanoparticles. *Journal of Materials Chemistry* **21**, 15121 (2011).
26. Goldman, E. B. *et al.* Biocompatibility of Tungsten Disulfide Inorganic Nanotubes and Fullerene-Like Nanoparticles with Salivary Gland Cells. *Tissue Engineering Part A* **21**, 1013–1023 (2015).
27. Pardo, M., Shuster-Meiseles, T., Levin-Zaidman, S., Rudich, A. & Rudich, Y. Low cytotoxicity of inorganic nanotubes and fullerene-like nanostructures in human bronchial epithelial cells : relation to inflammatory gene induction and antioxidant response. *Environmental Science and Technology* **48**, 3457–3466 (2014).
28. Ramachandran, K. *et al.* Tube Expansion Deformation Enables In Situ Synchrotron X-ray Scattering Measurements during Extensional Flow-Induced Crystallization of Poly l-Lactide Near the Glass Transition. *Polymers* **10**, (2018).
29. Ailianou, A., Ramachandran, K., Kossuth, M. B., Oberhauser, J. P. & Kornfield, J. A. Multiplicity of morphologies in poly (l-lactide) bioresorbable vascular scaffolds. *Proceedings of the National Academy of Sciences* **113**, 11670–11675 (2016).
30. Ramachandran, K. *et al.* Crimping-induced structural gradients explain the lasting strength of poly l-lactide bioresorbable vascular scaffolds during hydrolysis. *Proceedings of the National Academy of Sciences* **115**, 10239–10244 (2018).

31. Lalwani, G. *et al.* Tungsten disulfide nanotubes reinforced biodegradable polymers for bone tissue engineering. *Acta Biomaterialia* **9**, 8365–8373 (2013).
32. Bai, H. *et al.* Significantly improving oxygen barrier properties of polylactide via constructing parallel-aligned shish-kebab-like crystals with well-interlocked boundaries. *Biomacromolecules* **15**, 1507–1514 (2014).
33. Gorrasi, G. & Pantani, R. Hydrolysis and Biodegradation of Poly(lactic acid). in *Synthesis, Structure and Properties of Poly(lactic acid)* 119–151 (Springer, 2017).
34. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* **152**, 36–51 (2005).
35. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology* **116**, 71–6 (1996).
36. Mastronarde, D. N. Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. *Journal of Microscopy* **230**, 212–217 (2008).
37. Kumaraswamy, G., Verma, R. K. & Kornfield, J. a. Novel flow apparatus for investigating shear-enhanced crystallization and structure development in semicrystalline polymers. *Review of Scientific Instruments* **70**, 2097–2104 (1999).
38. Lilly, B. We Have Contact: Endothelial Cell-Smooth Muscle Cell Interactions. *Physiology* **29**, 234–241 (2014).
39. Foglia, S. *et al.* In vitro biocompatibility study of sub-5 nm silica-coated magnetic iron oxide fluorescent nanoparticles for potential biomedical application. *Scientific Reports* **7**, 1–13 (2017).
40. Appel, J. H. *et al.* Low Cytotoxicity and Genotoxicity of Two-Dimensional MoS_2 and WS_2. *ACS Biomaterials Science & Engineering* **2**, 361–367 (2016).
41. Janeschitz-Kriegl, H. *Poylmer Melt Rheology and Flow Birefringence*. (Springer-Verlag, 1983). doi:10.1007/978-3-642-68822-5
42. Sun, J. *et al.* Cytotoxicity, permeability, and inflammation of metal oxide nanoparticles in human cardiac microvascular endothelial cells. *Cell Biology and Toxicology* **27**, 333–342 (2011).
43. Lai, J. C. *et al.* Exposure to titanium dioxide and other metallic oxide nanoparticles induces cytotoxicity on human neural cells and fi broblasts. *International Journal of Nanomedicine* **3**, 533–545 (2008).
44. Fernandez-Ballester, L., Thurman, D. W. & Kornfield, J. A. Real-time depth sectioning: Isolating the effect of stress on structure development in pressure-driven flow. *Journal of Rheology* **53**, 1229 (2009).
45. Fernandez-Ballester, L., Thurman, D. W., Zhou, W. & Kornfield, J. A. Effect of long chains on the threshold stresses for flow-induced crystallization in iPP: Shish kebabs vs sausages. *Macromolecules* **45**, 6557–6570 (2012).
46. Kumaraswamy, G., Issaian, A. M. & Kornfield, J. A. Shear-enhanced crystallization in isotactic polypropylene. 1. Correspondence between in situ rheo-optics and ex situ structure determination. *Macromolecules* **32**, 7537–7547 (1999).

47. Kumaraswamy, G., Kornfield, J. A., Yeh, F. & Hsiao, B. S. Shear-enhanced crystallization in isotactic polypropylene. 3. Evidence for a kinetic pathway to nucleation. *Macromolecules* **35**, 1762–1769 (2002).
48. Kornfield, J. A., Kumaraswamy, G. & Issaian, A. M. Recent Advances in Understanding Flow Effects on Polymer Crystallization. *Industrial & Engineering Chemistry Research* **41**, 6383–6392 (2002).
49. Xu, H. *et al.* Formation of Shish-Kebabs in Injection-Molded Poly(L-Lactic acid) by Application of an Intense Flow Field. *Applied Materials & Interfaces* **4**, 6774–6784 (2012).
50. Sørensen, B. E. A revised Michel-Lévy interference colour chart based on first-principles calculations. *European Journal of Mineralogy* **25**, 5–10 (2013).
51. Kaplan-Ashiri, I. *et al.* On the mechanical behavior of WS2 nanotubes under axial tension and compression. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 523–528 (2006).
52. Samorodnitzky-Naveh, G. R., Redlich, M., Rapoport, L., Feldman, Y. & Tenne, R. Inorganic fullerene-like tungsten disulfide nanocoating for friction reduction of nickel–titanium alloys. *Nanomedecine* **4**, (2009).
53. Bojda, J. & Piorkowska, E. Shear-induced nonisothermal crystallization of two grades of PLA. *Polymer Testing* **50**, 172–181 (2016).
54. Xu, H., Xie, L. & Hakkarainen, M. Beyond a Model of Polymer Processing-Triggered Shear: Reconciling Shish-Kebab Formation and Control of Chain Degradation in Sheared Poly(l-lactic acid). *ACS Sustainable Chemistry and Engineering* **3**, 1443–1452 (2015).
55. Zhong, Y. *et al.* Rheologically determined critical shear rates for shear-induced nucleation rate enhancements of poly(lactic acid). *ACS Sustainable Chemistry and Engineering* **1**, 663–672 (2013).
56. Tang, H. *et al.* Shear flow and carbon nanotubes synergistically induced nonisothermal crystallization of poly(lactic acid) and its application in injection molding. *Biomacromolecules* **13**, 3858–3867 (2012).
57. Yamazaki, S., Itoh, M., Oka, T. & Kimura, K. Formation and morphology of 'shish-like' fibril crystals of aliphatic polyesters from the sheared melt. *European Polymer Journal* **46**, 58–68 (2010).

6.8 Supplementary Information

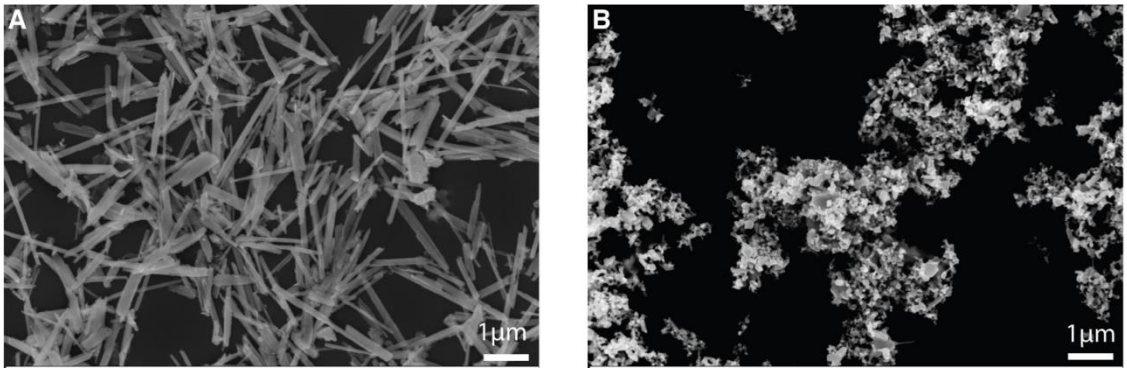

Figure S1. Scanning electron micrographs (SEM) of (**A**) Tungsten Disulfide (WS_2) nanotubes (WSNTs) and (**B**) Zinc Oxide (ZnO) nanoparticles.

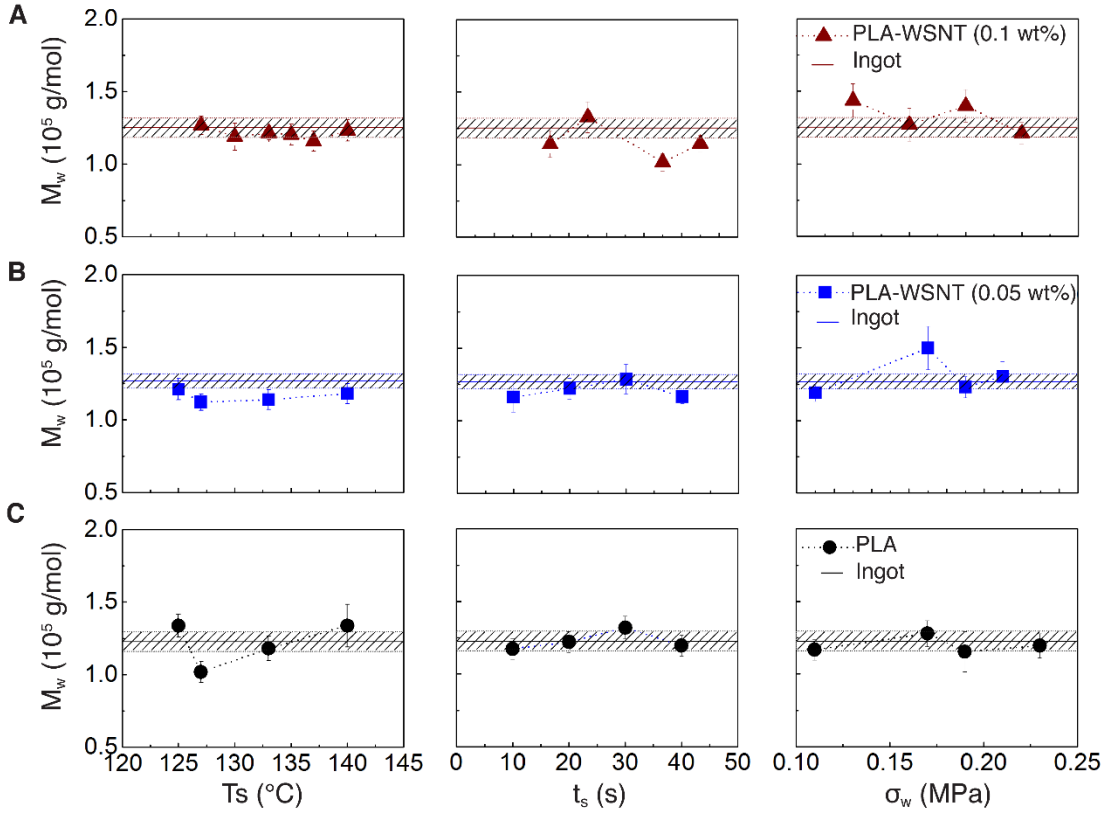

Figure S2. Weight-averaged molecular weights (M_w), inferred from the GPC traces in Figure S3, are presented for (**A**) PLA, (**B**) PLA-WSNT 0.05wt% and (**C**) PLA-WSNT 0.1wt% that were subjected to matched flow conditions under varying (*left*) shear temperature (Ts), (*middle*) shear duration (t_s) and (*right*) wall shear stress (σ_w). The M_w of the starting material for each experiment, labelled as "ingot", is presented as a horizontal line in each plot. The upper and lower horizontal lines indicate the uncertainty in the M_w of the ingot.

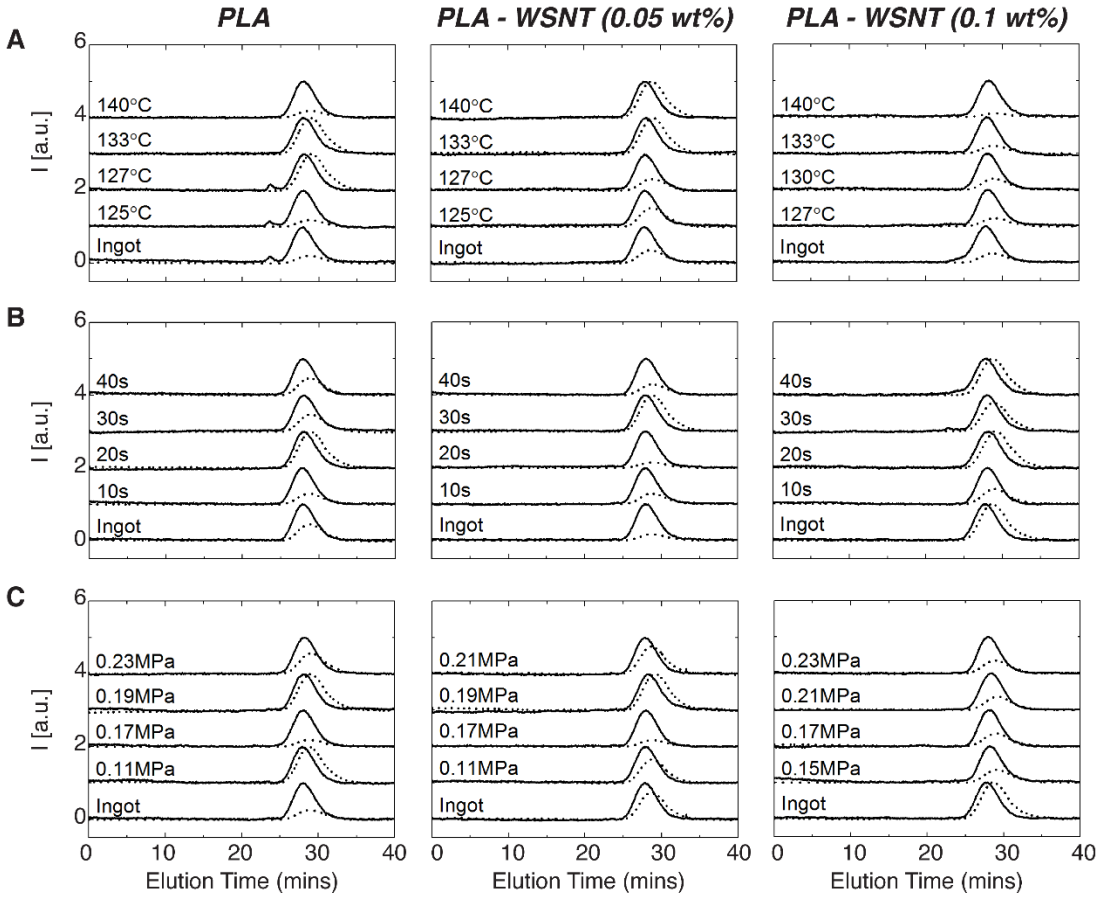

Figure S3 Selected gel-permeation chromatography (GPC) traces for extruded (*left*) PLA, (*mid*) PLA-WSNT (0.05wt%) and (*right*) PLA-WSNT (0.1 wt%) samples subjected to flow under varying (**A**) shear temperature (Ts), (**B**) shear duration (t_s) and (**C**) wall shear stress (σ_w). The selected traces encompass the start and the end of the short-term shear experiments (~4 hours) and show no sign of degradation when compared with the starting material (labelled as "ingot" in each plot). Traces from both the light scattering detector (solid lines)

and the refractive index detector (dashed lines) are presented. The molecular weights and GPC traces in Figures S2–S3 correspond to the *in situ* retardance profiles in Figure 4.

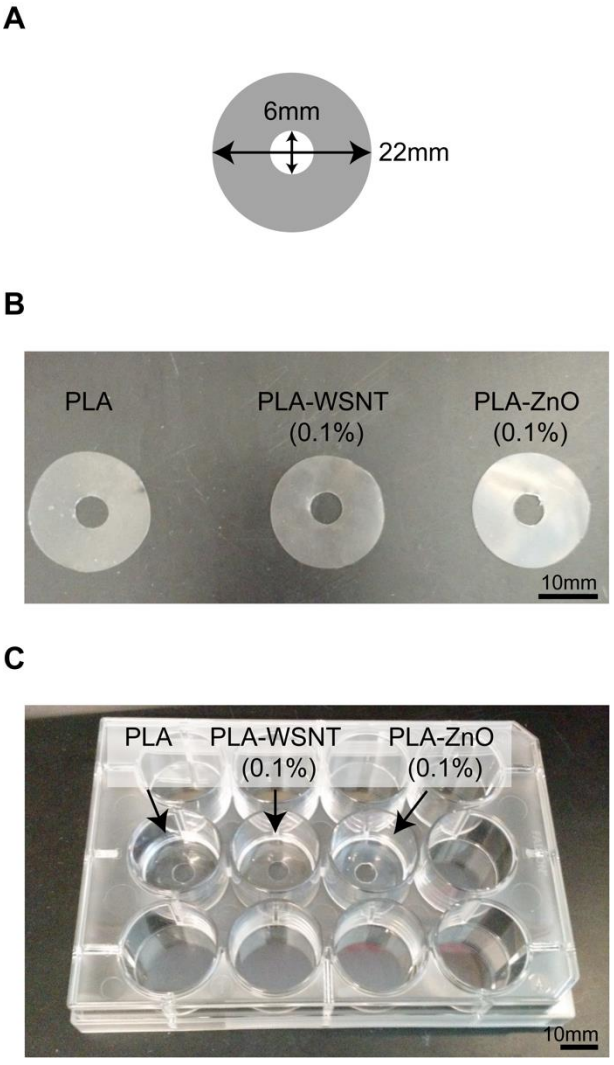

Figure S4. (**A**) Dimensions of the PLA/nanocomposite disks that are cut from solvent-cast films. (**B**) Photographs of PLA, PLA-WSNT 0.1 wt% and PLA-ZnO 0.1wt% disks. (**C**) Photograph of the disks in (B) submerged in 12-well plates containing media.

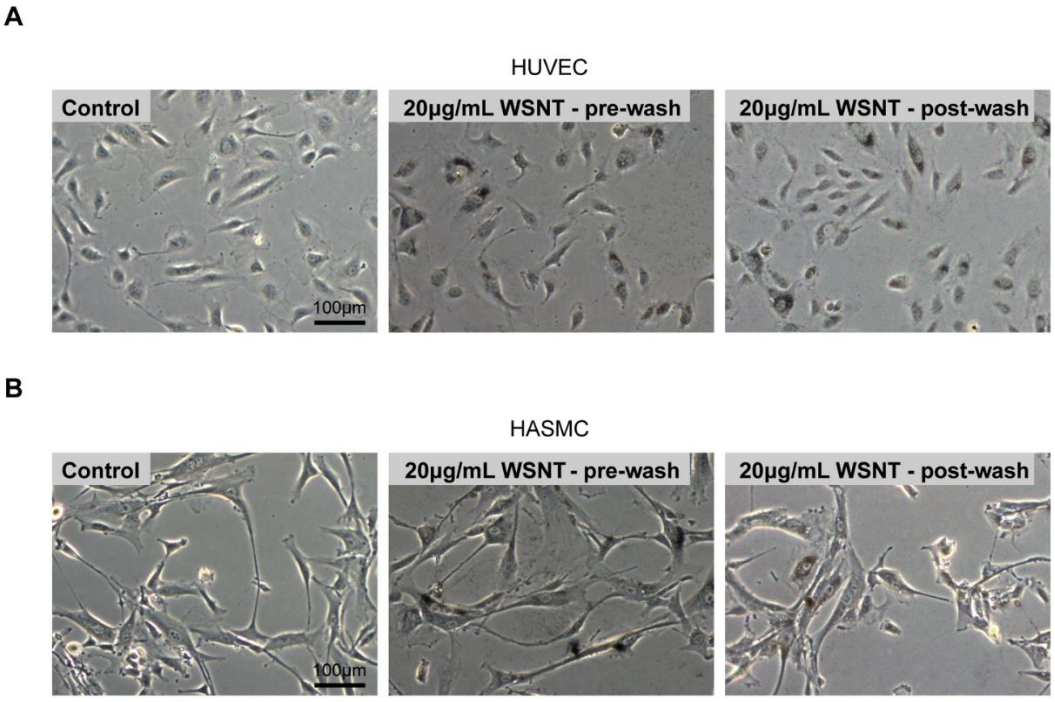

Figure S5. Phase contrast micrographs of (**A**) HUVECs and (**B**) HASMCs in (*left*) vehicle control wells, (*mid*) wells treated with 20µg/mL WSNT for 24 hours and (*right*) wells treated with 20µg/ml WSNTs that are subsequently washed after 24 hrs to remove non-endocytosed nanoparticles.

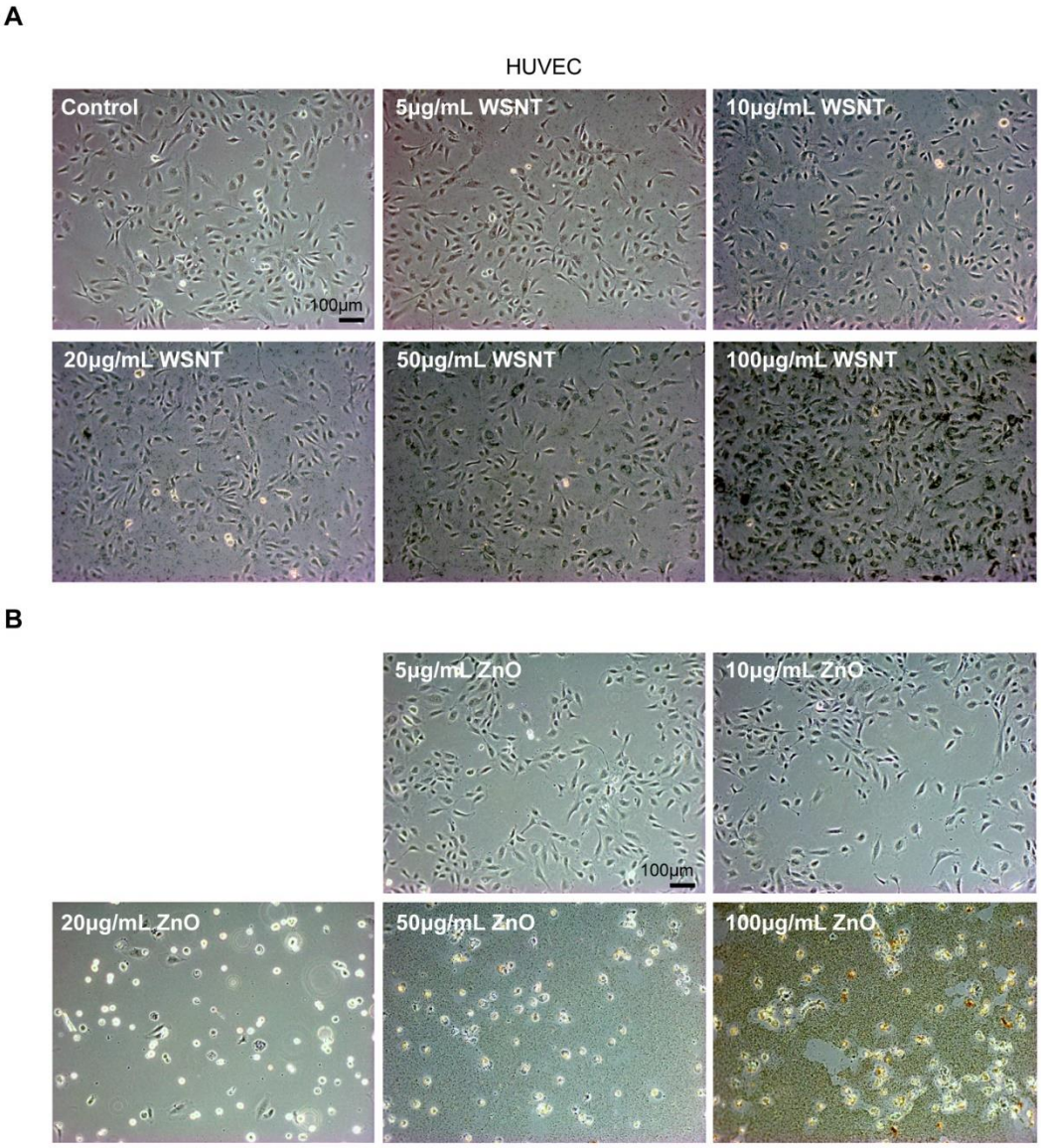

Figure S6. Phase contrast micrographs of HUVECs treated with increasing concentrations of (**A**) WSNTs and (**B**) ZnO nanoparticles for 24 hours.

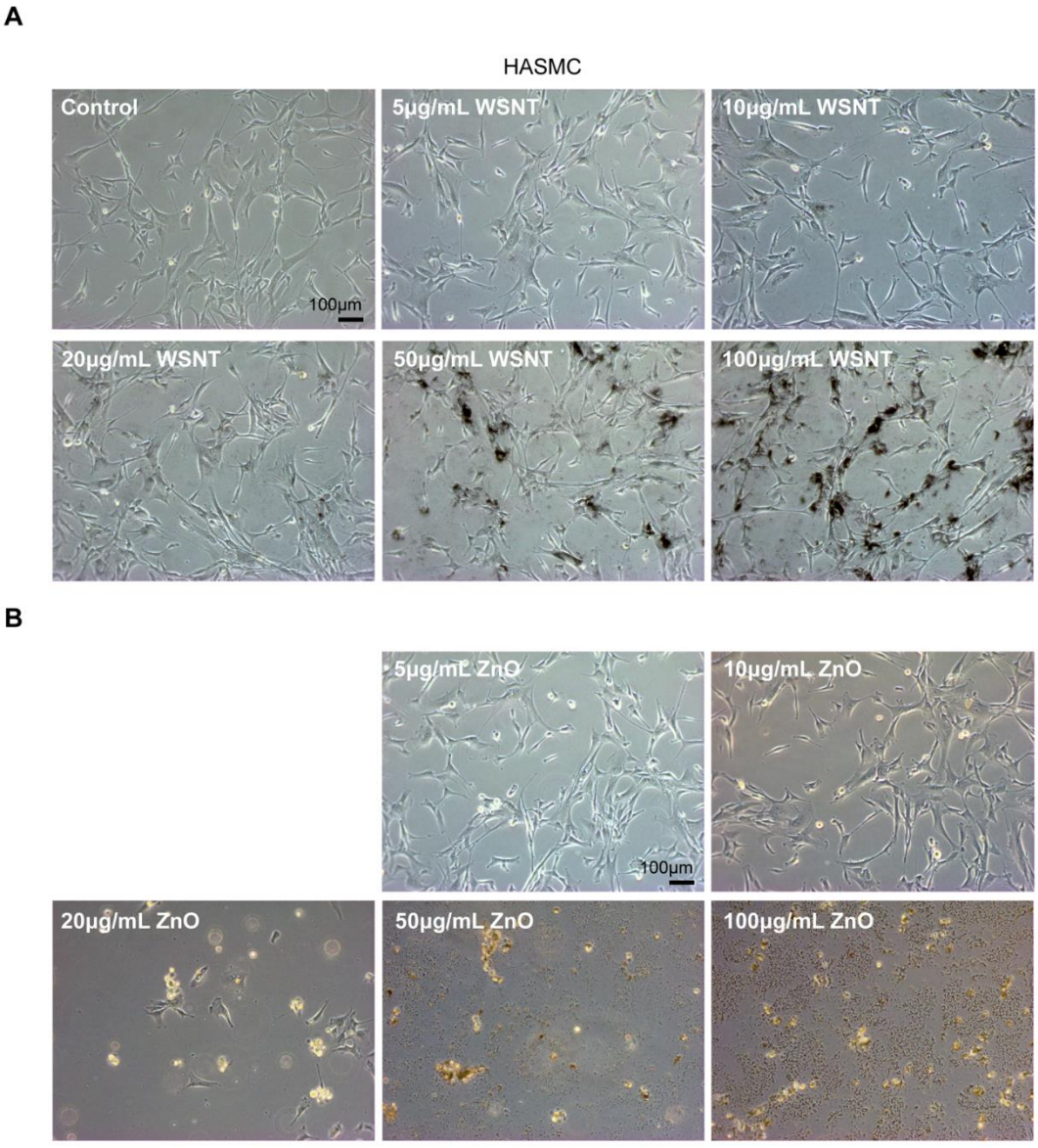

Figure S7. Phase contrast micrographs of HASMCs treated with increasing concentrations of (**A**) WSNTs and (**B**) ZnO nanoparticles for 24 hours.

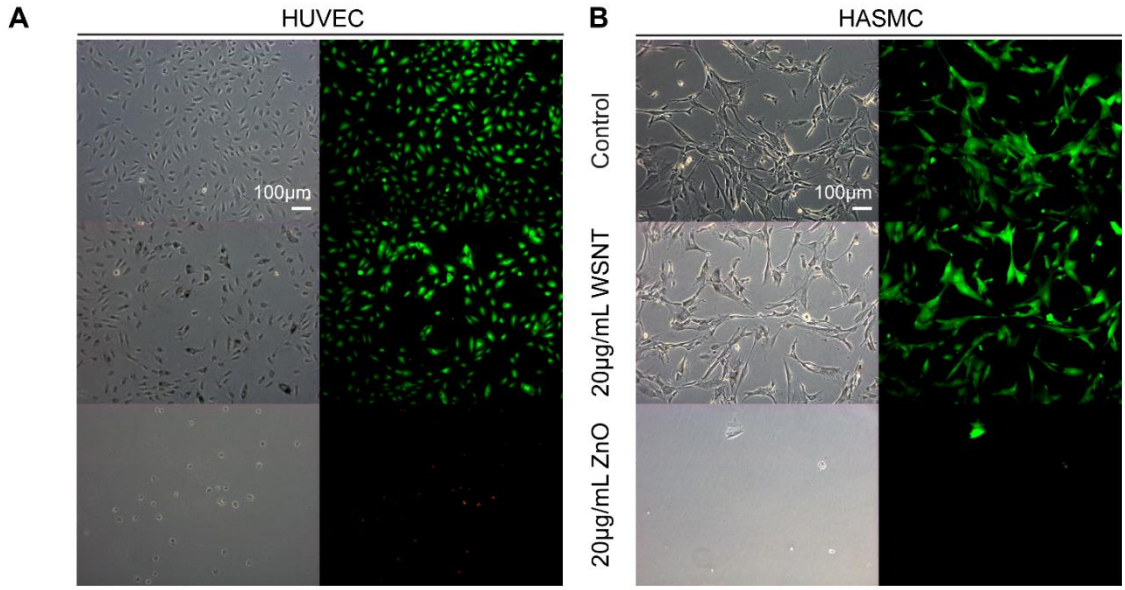

Figure S8. Live/Dead staining images of (**A**) HUVECs and (**B**) HASMCs after 24 hours of treatment (with and without bare nanoparticles). (A-B, *left*) Phase contrast micrographs and (A-B, *right*) merged dead (red) and live (green) staining images of the treated cells. A majority of the cells treated with ZnO underwent detachment and were subsequently removed during the assay wash steps.

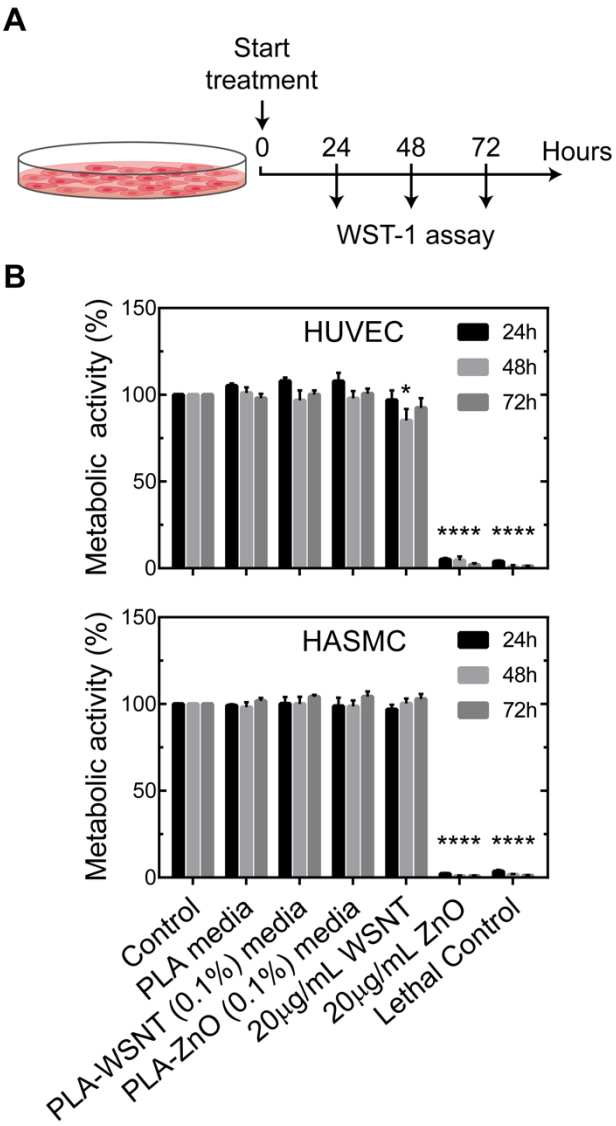

Figure S9. Time course cell viability assay for HUVECs and HASMCs exposed to bare nanoparticles and disk-treated media with appropriate controls. (**A**) Schematic of HUVECs and HASMCs grown in 96-well plates before the media is replaced with one that either contains nanoparticles or has been conditioned with polymer disks for 24 hours. Separate samples were used at 24, 48 and 72 hours after treatment for the WST-1 assay. Experiments were performed three times with six replicates per condition. (**B**) Metabolic activity measurements generated by the WST-1 assay. All statistical tests were performed against vehicle control samples at each respective time point. Mean and standard error of mean plotted. The stars indicate the magnitude of adjusted p-value (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$).

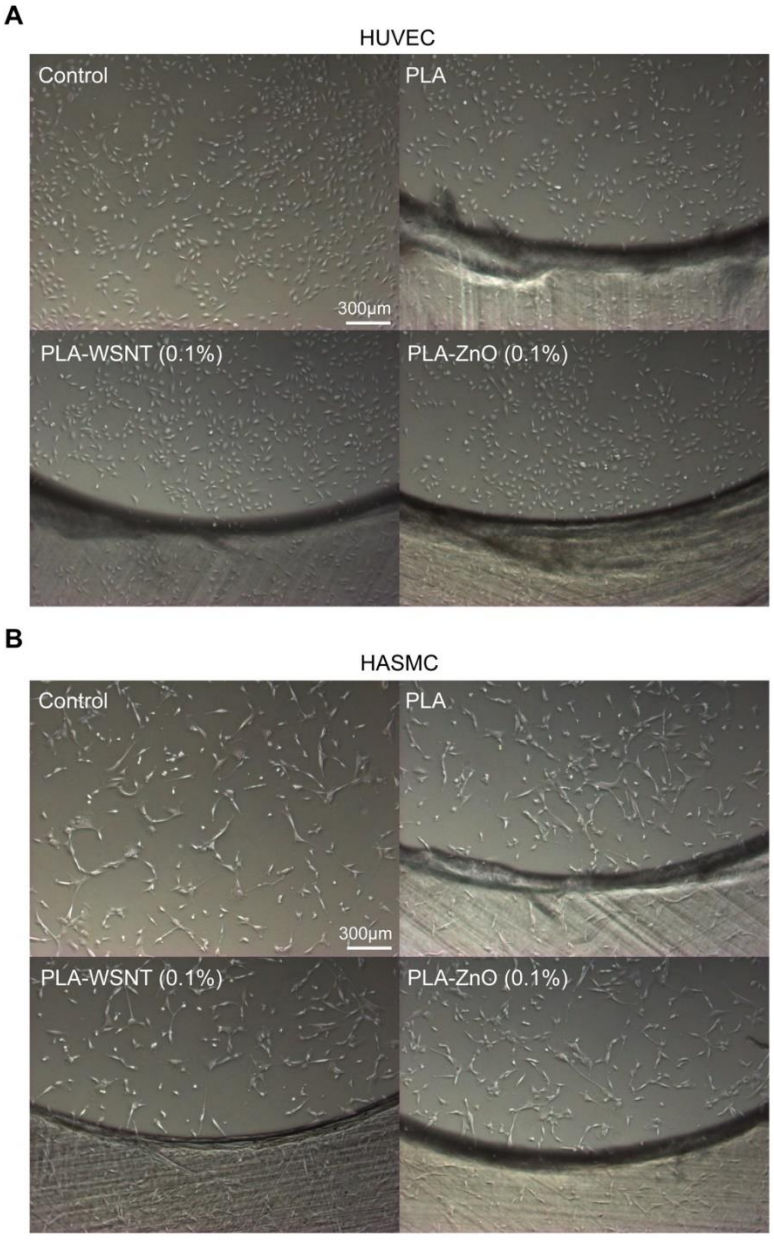

Figure S10. Phase contrast micrographs of (**A)** HUVECs and (**B**) HASMCs exposed to PLA, PLA-WSNT 0.1wt% and PLA-ZnO 0.1 wt% disks for 24 hours. The images indicate consistent cell density across the edge of the disk in each case.

PLA and PLA-WSNT are subjected to matched thermal histories during shear experiments

We probe the influence of WSNTs on the flow-induced crystallization of PLA by systematically varying the crystallization temperature (nucleant effects increase with increasing temperature above the fastest-crystal growth temperature), shearing time (nuclei formed by PLA increase with increasing shear time, but the number of nanoparticles does not), and shear stress (which correlates with different degrees of polymer orientation in the melt). We start by erasing thermal history by heating the sample to 200°C. In every case, the sample is subjected to the same thermal profile during cooling (Figure S11, *left*, time zero set to the onset of cooling). When the temperature at the observation point ("Outlet") reaches the desired shearing temperature, short term shear is induced by rapidly imposing a high pressure at the inlet of the rectangular channel (Figure S11, *right*, time zero set to 0.02s prior to the onset of the pressure rise) and rapidly releasing the inlet pressure to stop shearing when the desired shear time (t_s) is reached. Note that we vary the temperature during the shear pulse, but impose the same growth conditions by selecting the moment at which the shear pulse is triggered (Figure S11A, *left*, gray and black lines show the moment shear is triggered for Ts = 140°C and 125°C, respectively, since the moment cooling began; Fig S11A, *right*, shows seven 10s pressure pulses with respect to time from the onset of the shear pulse). For the remaining protocols, the shear pulse was triggered at the moment the temperature falls to 130°C to examine the effects of either shear time (t_s, Figure S11B) or wall shear stress (σ_w, proportional to the pressure drop from inlet to exit of the channel shown in Figure S11C)

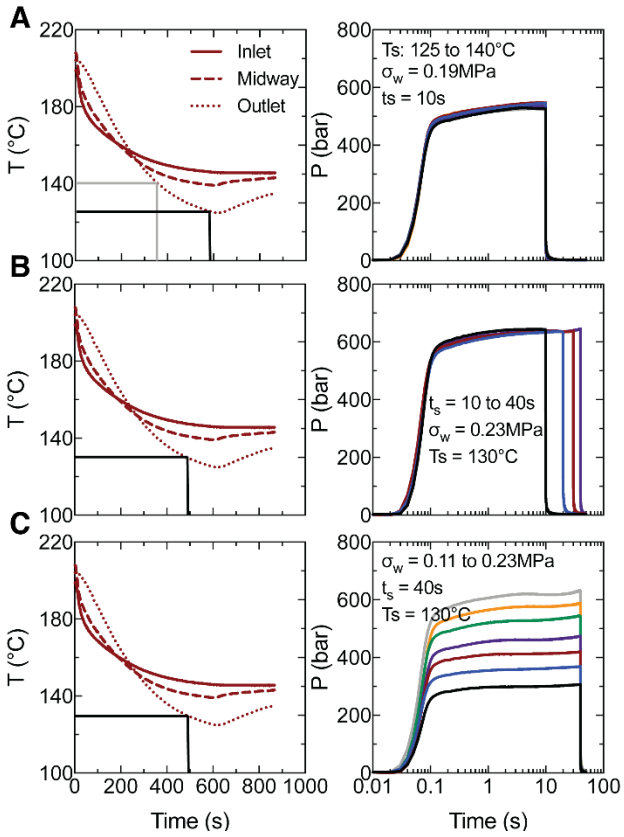

Figure S11. Temperature (*left*) and pressure traces (*right*) for short-term shear experiments with varying (**A**) shear temperature (Ts), (**B**) shear duration (t_s), and (**C**) wall shear stress (σ_w). The temperature of the flow cell is measured near the inlet, at the outlet and midway between the inlet and the outlet. The flow-cell is equipped with a pair of quartz windows at the outlet to permit acquisition of *in situ* retardance data. Therefore, the temperature at the outlet is used to trigger shear at a prescribed Ts. The inlet is deliberately kept ~20°C warmer than the outlet at the time of shear to prevent choking at the inlet. A pressure transducer records the pressure imposed at the inlet of the flow-cell and is also used to infer the wall shear stress experienced by the sample. In (A), Ts is varied from 140°C (gray lines) to 125°C (black lines) using the same shear duration (10s) and wall shear stress (0.19 MPa). In (B) and (C), Ts is held constant (130°C, black lines) but the shear duration (t_s: 10–40s) and wall shear stress (σ_w: 0.11–0.23 MPa) are varied respectively. In (A-C, *left*), the temperature is plotted with respect to the onset of cooling. In (A-C, *right*), the pressure traces are plotted with respect to the onset of shear at a specific Ts. The thermal and deformation histories presented in (A-C) correspond to the *in situ* retardance data presented in Figure 4A-C and Figure S13–15.

Calculation of the wall shear rate

In the experiments described in this report, we impose flow at a fixed wall shear stress (σ_w), which is inferred from the pressure drop (ΔP), and the width (w: 0.5mm), height (h: 6.35mm) and length (l: 63.5mm) of the channel (*S1*)[1]. By shearing at a prescribed wall shear stress, we allow the material to respond to the imposed flow at a shear rate that reflects its viscosity and M_w distribution.

$$\sigma_w = \frac{\Delta P}{l}\left(\frac{h}{2\,(1 + h/w)}\right) \qquad [S1]$$

$$Q = \frac{mass\ extruded}{shear\ duration\ .\ \rho_{PLA}} \qquad [S2]$$

$$\dot{\gamma_a} = \frac{6Q}{w.h^2} \qquad [S3]$$

Using the volumetric flow rate (Q), which is calculated from the density of PLA (ρ : 1.25 mg/mm^3)[2] and the mass of extruded material (*S2*), we compute an apparent shear rate that is applicable to Newtonian flow (*S3*). In order to account for the non-Newtonian behavior of polymer melts, we apply a correction (*S4*) based on the Weissenberg–Rabinowitsch–Mooney equation[3]. However, this correction requires measurement of the apparent shear rate as a function of the wall shear stress. Thus, Figure S12 reports the apparent wall shear rate $\dot{\gamma_a}$ for experiments performed at constant σ_w (Figure 4A-B) and the true wall shear rate $\dot{\gamma_w}$ for experiments performed with varying σ_w (Figure 4C).

$$\dot{\gamma}_w = \frac{\dot{\gamma}_a}{3}\left[2 + \frac{d(\ln \dot{\gamma}_a)}{d\,(\ln \sigma_w)}\right] \qquad [S4]$$

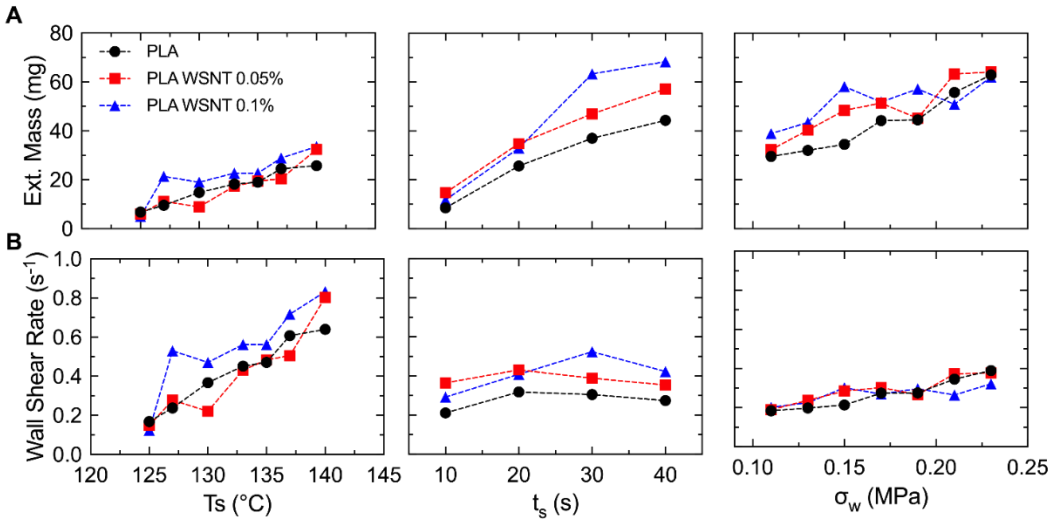

Figure S12. (**A**) The mass extruded from each short-term shear experiment is used to compute the corresponding (**B**) wall shear rate $\dot{\gamma}_w$ for data acquired at varying (*left*) shear temperature (Ts), (*mid*) shear duration (t_s) and (*right*) wall shear stress (σ_w). The apparent wall shear rate is presented in (B, *left–mid*); the true wall shear rate is only presented in (B, *right*) as the correction requires the variation in the mass extruded with respect to the wall shear stress.

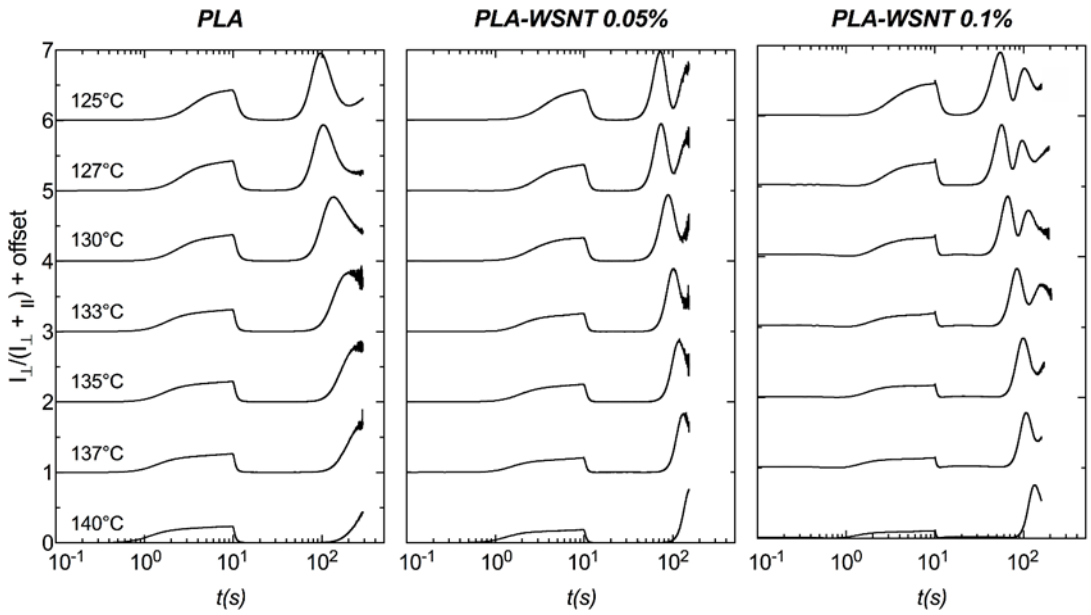

Figure S13. Normalized intensity profiles for (*left*) PLA, (*mid*) PLA-WSNT 0.05 wt% and (*right*) PLA-WSNT 0.1 wt% at varying shear temperature (Ts: 125–140°C). The samples were extruded for a duration (t_s) of 10s at a wall shear stress (σ_w) of 0.19 MPa.

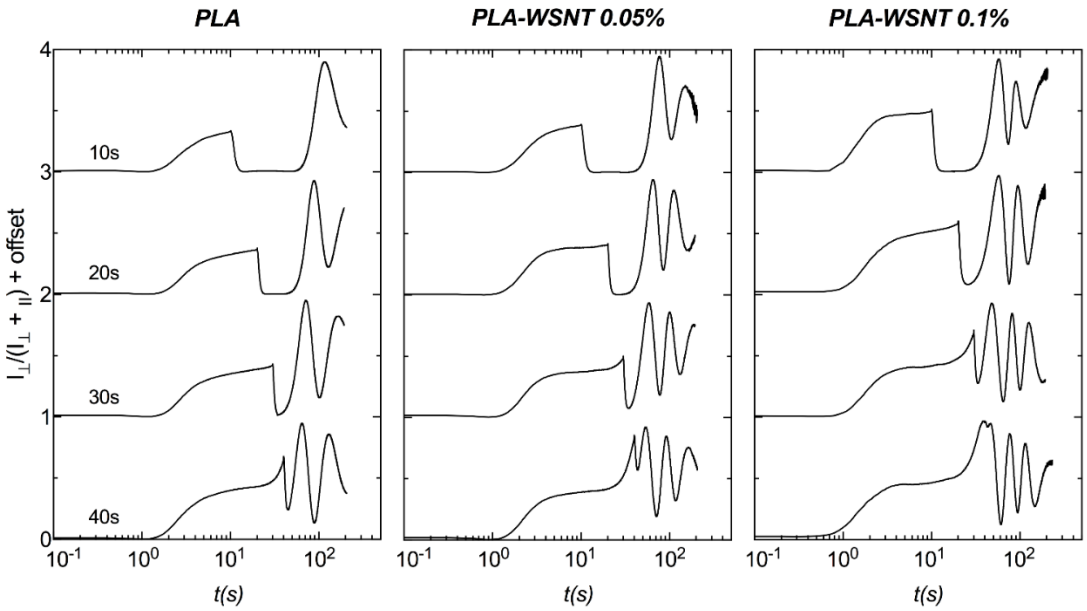

Figure S14. Normalized intensity profiles for (*left*) PLA, (*mid*) PLA-WSNT 0.05 wt% and (*right*) PLA-WSNT 0.1 wt% at varying shear duration (t_s: 10–40s). The samples were extruded at Ts = 130°C at a wall shear stress (σ_w) of 0.23 MPa.

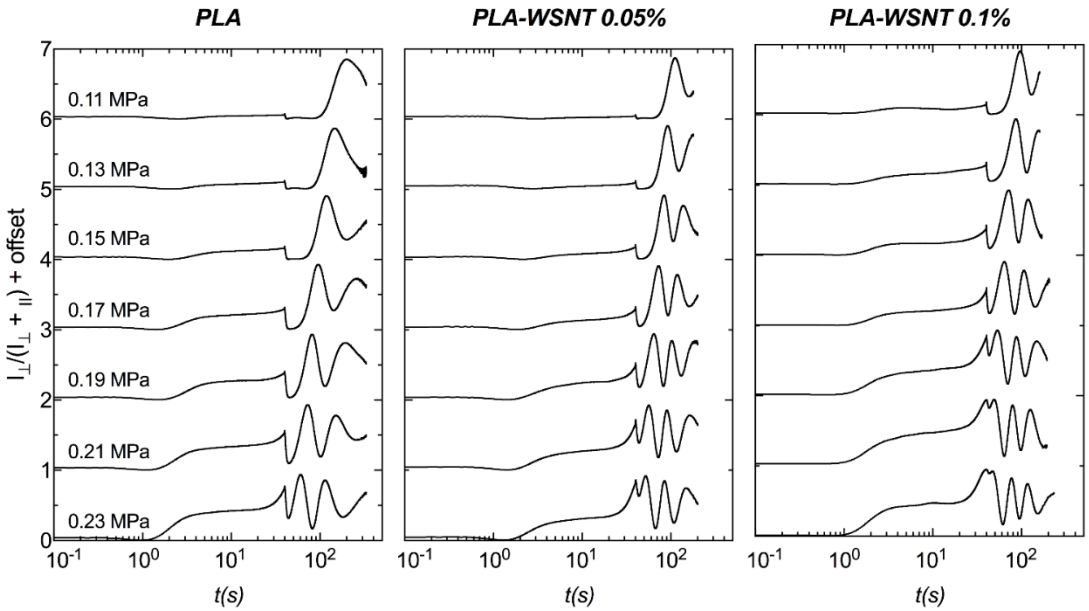

Figure S15. Normalized intensity profiles for (*left*) PLA, (*mid*) PLA-WSNT 0.05 wt% and (*right*) PLA-WSNT 0.1 wt% at varying wall shear stress (σ_w: 0.11–0.23 MPa). The samples were extruded at Ts = 130°C for a duration (t_s) of 40s.

Estimating the average spacing between WSNTs dispersed in the PLA matrix

Let us consider a 1cm^3 volume element of PLA that contains 0.1 wt% of homogenously dispersed WSNTs (diameter ~200nm, length ~ 2μm). Based on the density of PLA (1.25 g/cm^3) and WS_2 (7.5 g/cm^3), this volume of element is expected to contain 1.3 μg of WS_2 that occupies a volume of ~1.73 x 10^7 μm^3. As a result, 0.1 wt% of WSNTs occupy 0.0017 volume % in the PLA matrix.

Based on the volume of an individual nanotube (0.0628 μm^3), 1 cm^3 of PLA, which contains 1.73 x 10^7 μm^3 WS_2, is expected to contain ~ 2.75 x 10^8 nanotubes. Therefore, the volume of polymer surrounding each individual nanotube is ~ 36μm^3. Assuming uniform dispersion, the average spacing between WSNTs is the cube root of 36μm^3 or ~ 3.3 μm.

Calculating the birefringence associated with WSNTs

To determine the birefringence associated with 0.1 wt% WSNTs relative to the PLA melt, we apply Weiner's equations[4] that were developed for an analogous case of microtubules dispersed in a medium. The birefringence (*Δn, S5*) is defined as the difference between the extraordinary (n_e) and ordinary (n_o) refractive indices.

$$\Delta n = n_e - n_o \qquad [S5]$$

We compute n_e (*S6*)[4] and n_o (*S7*)[4] using refractive indices of WSNTs (n_1 = 4)[5] and PLA (n_2 = 1.46)[6], and the volume fraction of WSNTs (f = 1.7x10^{-5} for 0.1wt% WSNTs) in the PLA melt.

$$n_e = n_2 \sqrt{1 + \frac{(n_1^2 - n_2^2)}{n_2^2}.f} \qquad [S6]$$

$$n_o = n_2 \sqrt{\frac{n_1^2 + n_2^2 + f(n_1^2 - n_2^2)}{n_1^2 + n_2^2 - f(n_1^2 - n_2^2)}} \qquad [S7]$$

From equations S5–S7, we infer that the birefringence associated with 0.1wt% of WSNTs is ~ 6.17×10^{-5}, which translates to a retardance (δ, *S8*) of ~0.3 for the optical train used in our instrument [sample thickness (*d*): 500μm and laser wavelength (λ): 632nm]. At 0.05wt% WSNTs (0.00085 vol%), the resulting retardance is ~0.15 (for $f << 1$, the birefringence is proportional to *f*).

$$\delta = \frac{2\,\pi\,\Delta n\,d}{\lambda} \qquad [S8]$$

Depth-sectioning the *in situ* PLA and PLA-WSNT retardance traces

The depth-sectioning approach was developed by Fernandez-Ballester et al[7] to isolate the retardance contributed by individual layers of material through the thickness of isotactic polypropylene subjected to a shear flow. We apply the same approach to our experiments with PLA and PLA-WSNT. We begin by acquiring *in situ* retardance data at the highest wall shear stress ($\sigma_{w,i}$), and then collect data at progressively lower wall shear stresses ($\sigma_{w,i+1}$).

$$\Delta\delta\left(\sigma_{w,i},\sigma_{w,i+1}\right)_{app} = \delta\left(\sigma_{w,i}\right) - \frac{\sigma_{w,i+1}}{\sigma_{w,i}}\,\delta\left(\sigma_{w,i+1}\right) \qquad [S9]$$

$$\Delta\delta\left(\sigma_{w,i},\sigma_{w,i+1}\right)_{true} = \frac{\sigma_{w,i}}{\left(\sigma_{w,i} - \sigma_{w,i+1}\right)} \cdot \frac{1}{d/_2} \cdot \Delta\delta\left(\sigma_{w,i},\sigma_{w,i+1}\right)_{app} \qquad [S10]$$

The depth-sectioning approach relies on the linear variation of shear stress from the walls (maximum) to the core (zero) of the sample (the thickness, *d*, is 500μm along the velocity gradient direction). It enables isolation of the residual or depth-sectioned retardance contributed by a slice of material that experiences a shear stress varying from ($\sigma_{w,i}$) to ($\sigma_{w,i+1}$) by subtraction of the retardance profiles acquired at the two different wall shear stresses. Prior to subtraction, the data acquired at the lower wall shear stress $\delta(\sigma_{w,i+1})$ is rescaled by the ratio of the lower ($\sigma_{w,i+1}$) to the higher wall shear stress ($\sigma_{w,i+1}$) (*S9*). The resulting residual retardance is then normalized by the thickness of the layer, which is calculated based on the linear shear stress profile (see correction factor before the residual retardance term, $\Delta\delta$, in *S10*).

In Figure 7, depth-sectioned retardance profiles are presented for layers located ~20μm, 40μm, 65μm and 85μm from the wall of the channel. Table S1 below lists the pair of retardance profiles from Figure 4C that were used to compute the depth-sectioned profiles in Figure 7.

Table S1: Pairs of retardance traces acquired at differing wall shear stresses that were used to compute the depth-sectioned traces

Layer Position	Residual Retardance
~20μm	$\Delta\delta$ ($\sigma_{0.21MPa}$, $\sigma_{0.17MPa}$)
~40μm	$\Delta\delta$ ($\sigma_{0.19MPa}$, $\sigma_{0.15MPa}$)
~65μm	$\Delta\delta$ ($\sigma_{0.17MPa}$, $\sigma_{0.13MPa}$)
~85μm	$\Delta\delta$ ($\sigma_{0.15MPa}$, $\sigma_{0.11MPa}$)

Quantifying the upturn and residual retardance

We present select retardance traces for PLA-WSNT (0.05 wt%) acquired at three different wall shear stresses (σ_w: 0.11, 0.17, and 0.23 MPa) to illustrate how we compute the height of the upturn (Figure S16) and the residual retardance (Figure S17). The same method is applied to PLA and PLA-WSNT (0.1 wt%) data.

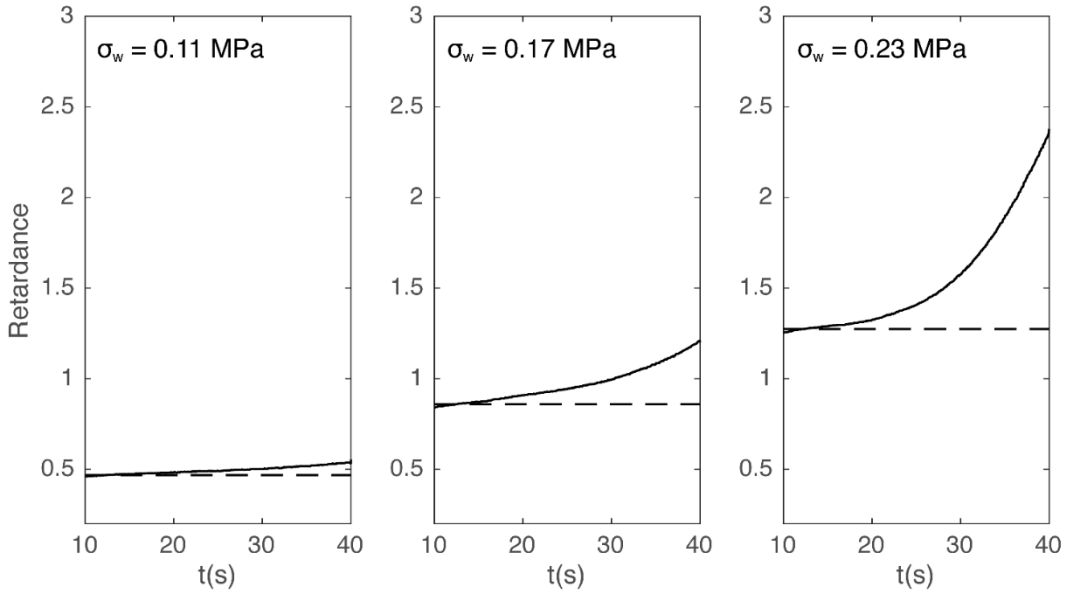

Figure S16. Retardance traces acquired at a wall shear stress of (*left*) 0.11MPa, (*mid*) 0.17MPa and (*right*) 0.23MPa are plotted for $10 \leq t(s) \leq 40s$ to illustrate the increase in the upturn as a function of increasing wall shear stress. To quantify the height of the upturn, we subtract a baseline (dashed black line) that corresponds to the plateau retardance (taken as the average of all values for $10 \leq t(s) \leq 15$) from the retardance at the end of the shear pulse ($t = 40s$).

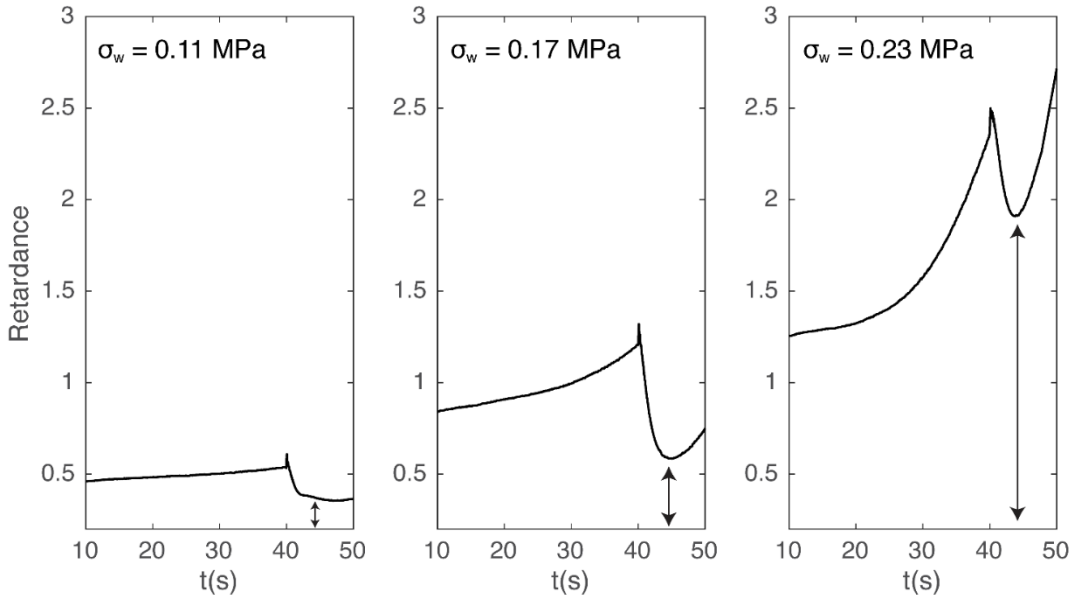

Figure S17. Retardance traces acquired at a wall shear stress of (*left*) 0.11MPa, (*mid*) 0.17MPa and (*right*) 0.23MPa are plotted for $10 \leq t(s) \leq 50$s to illustrate the correspondence between the height of the upturn (see Figure S16) and the residual retardance after cessation of flow. The black arrows (identified as the minima in retardance traces for $40 \leq t(s) \leq 50$) quantify the increase in the residual retardance as a function of increasing wall shear stress.

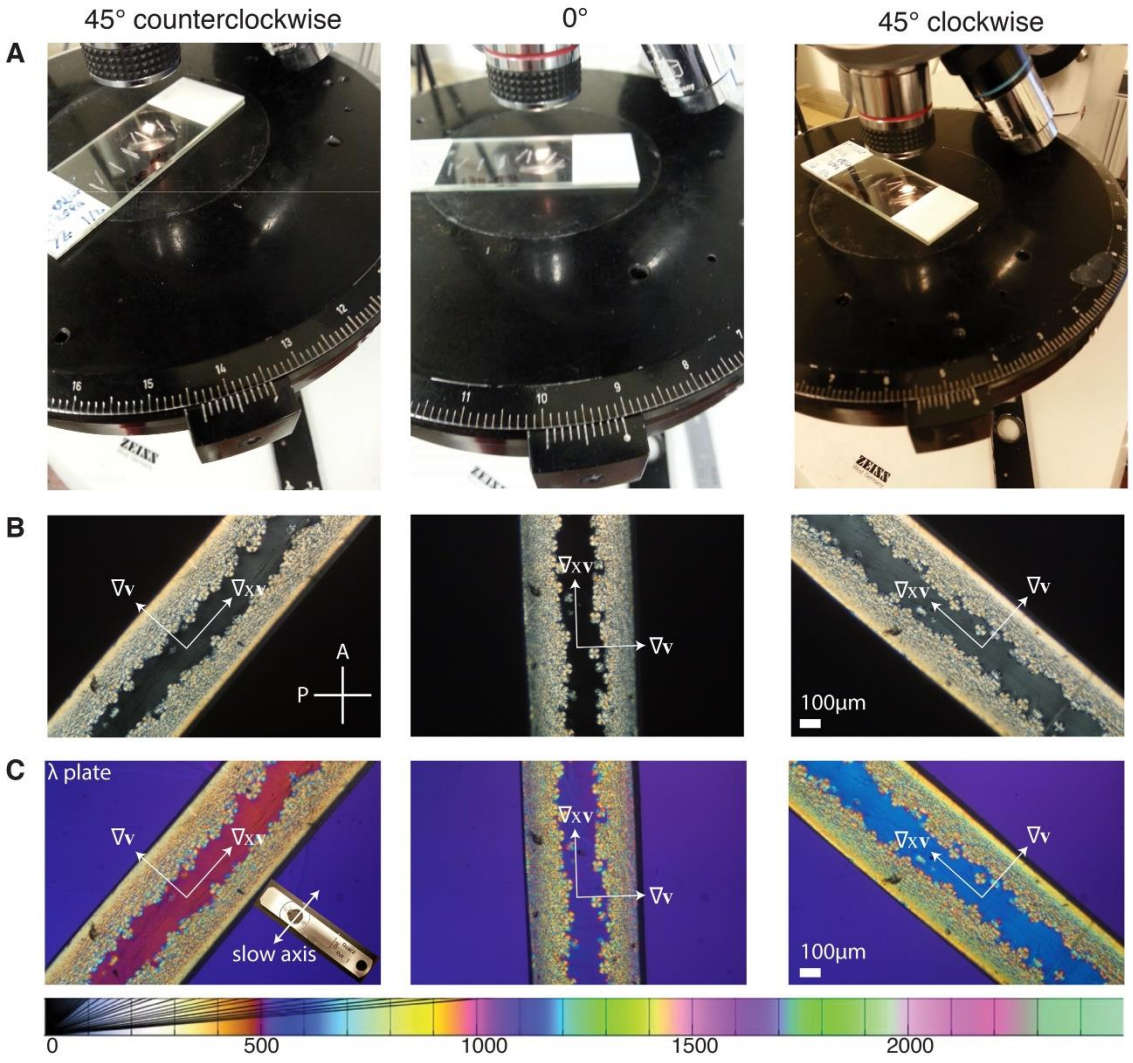

Figure S18. PLA sections (σ_w = 0.23MPa, Ts = 130°C and t_s = 40s; ∇x**v**–∇**v** plane) are rotated in the microscope in 45° intervals (left to right) to highlight anisotropy in the core and the skin. The same approach is applied to PLA-WSNT sections as well (see Figure S19). (**A**) Photographs of microtomed PLA sections (~50µm thick) in the microscope and (**B-C**) polarized light micrographs of a PLA section in all three orientations (-45°, 0° and 45°) (**B**) without a retardation plate and (**C**) with a full-wave retardation plate. The orientation of the analyzer and polarizer for all images in (B-C) is indicated in (B, left). The orientation of the slow axis of the full-wave retardation plate for the images in (C) is indicated in (C, left). The Michel-Levy color chart (adapted from Ref. [50] of main text) at the bottom connects the observed colors to the retardation in nm.

Polarized light microscopy indicates oriented crystallites normal to the flow direction

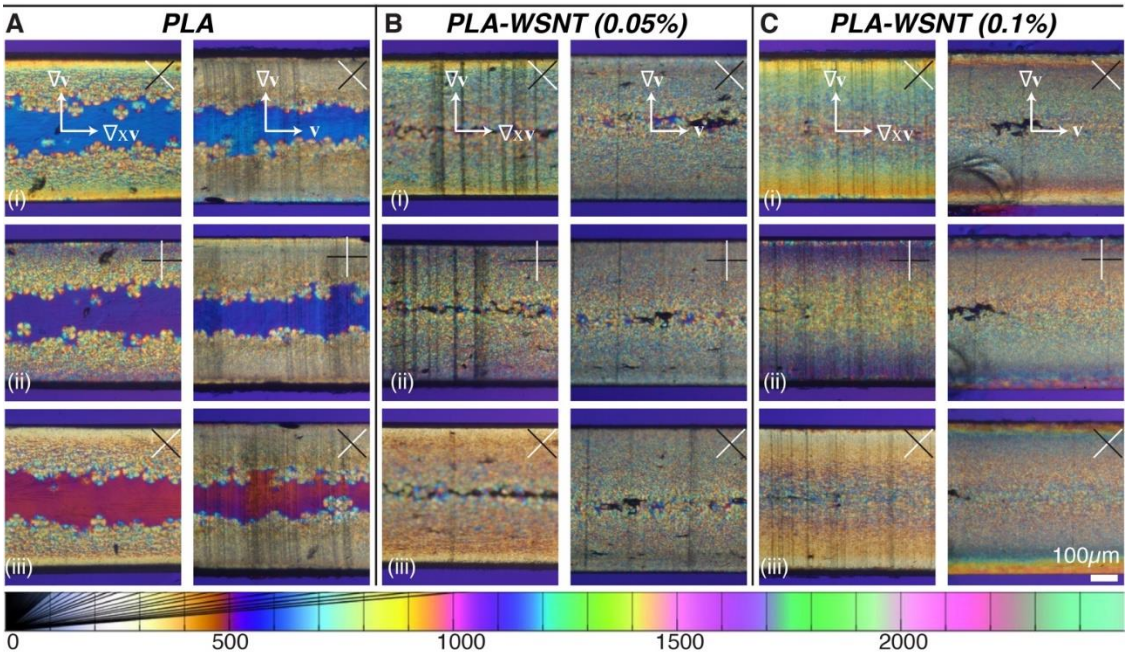

Figure S19. Polarized light micrographs acquired at three different orientations of the ~50μm thick sections relative to the polarizer and the analyzer (indicated in the top-right corner of each image) from samples sheared for t_s = 40s at T_s = 130°C and σ_w = 0.23MPa for (**A**) PLA, (**B**) PLA-WSNT 0.05wt% and (**C**) PLA-WSNT 0.1wt%. Images were acquired through linear crossed polarizers with a full-wave retardation plate (see Figure S18 for details). Pairs of images show sections cut in the vorticity–velocity gradient plane ($\nabla x\mathbf{v}$ –$\nabla\mathbf{v}$) on the left and flow–velocity gradient plane ($\mathbf{v}$ –$\nabla\mathbf{v}$) on the right). The top row of images are the same t_s = 40s images in Fig 6A–C; from top to bottom are successive 45° rotations of the sample. Note that the slow axis of the waveplate is vertical in the top image, CW 45° in the middle image and horizontal in the bottom image. The Michel-Levy color chart (adapted from Ref. [50] of main text) at the bottom connects the observed colors to the retardation in nm.

Wide Angle X-ray scattering reveals orientation of crystallites in PLA

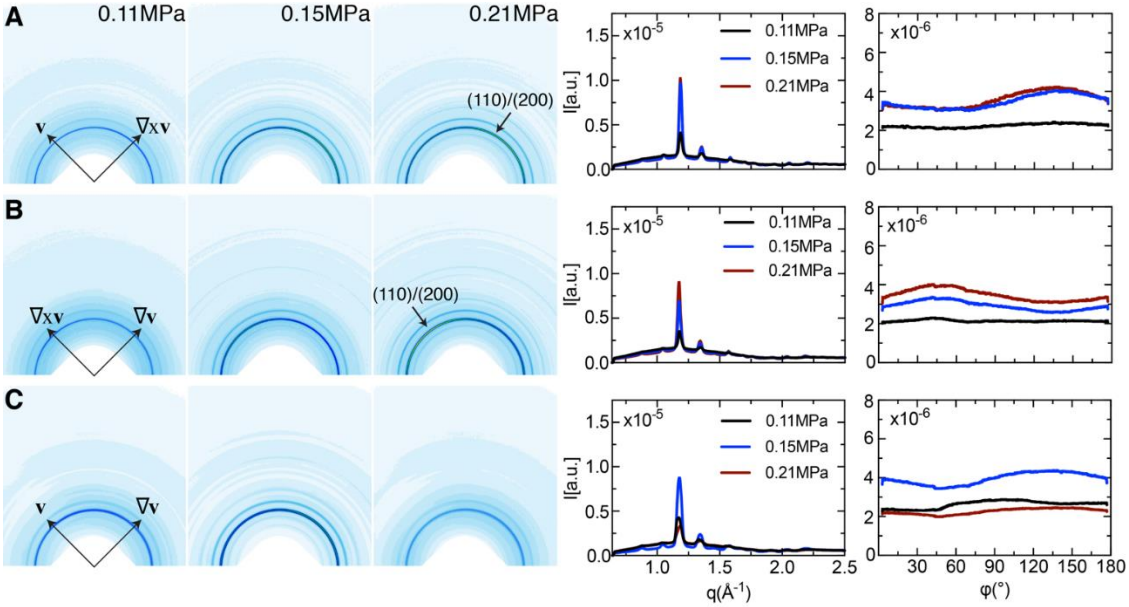

Figure S20. Wide angle X-ray scattering (WAXS) data were acquired on PLA sections subjected to varying wall shear stress (σ_w = 0.11, 0.15 and 0.21MPa at Ts = 130°C and t_s = 40s) along the (**A**) flow – vorticity plane (**v** – ∇x**v**), (**B**) vorticity – velocity gradient plane (∇x**v** – ∇**v**) and (**C**) flow – velocity gradient plane (**v** – ∇**v**). WAXS data are presented as (A-C, *left*) 2D patterns and (A-C, *right*) azimuthally [I(*q*)] and radially averaged [I(*φ*)] plots. Radial averaging for the I(*φ*) plots is performed in the vicinity of the (110)/(200) diffraction: $1.13 \leq q \leq 1.23$ Å^{-1}. The WAXS patterns are presented using a custom colormap with a scale that varies from 0 [white] to 1.75e^{-5} [red] counts. The crystallinity observed for PLA in the flow – gradient plane (part C) is higher than usual as the samples have been subjected to a thermal ramp (heat from 30°C to 180°C at 10°C/min and immediately cool back to 30°C at 10°C/min) prior to data acquisition. Transient X-ray patterns acquired during the ramp indicate that the material gains crystallinity and preserves the orientation distribution of crystallites.

Wide Angle X-ray scattering reveals orientation of crystallites in PLA-WSNT 0.05wt%

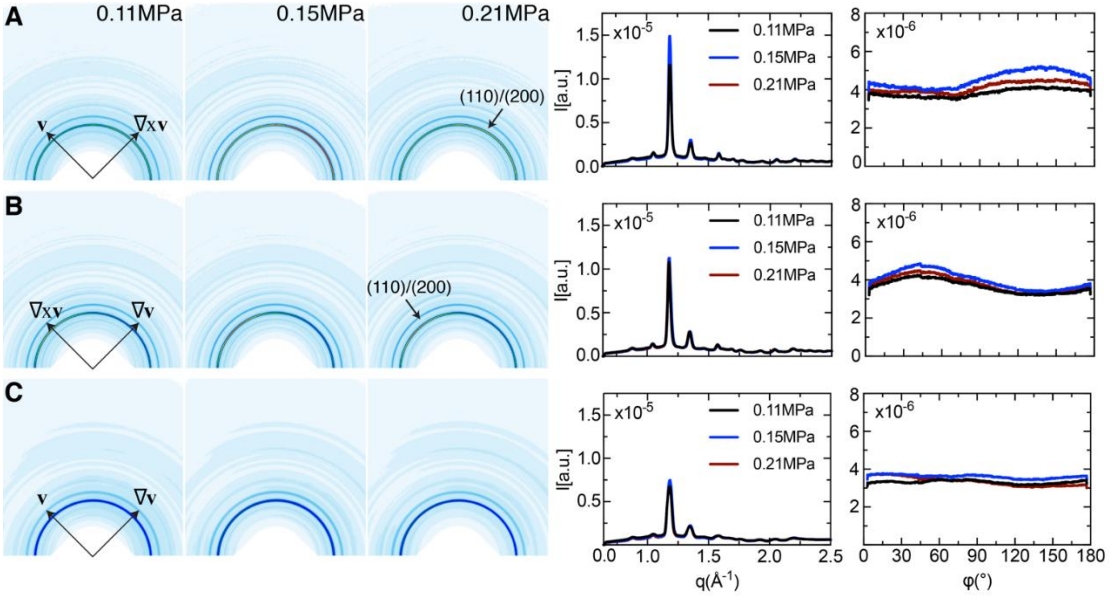

Figure S21. Wide angle X-ray scattering (WAXS) data were acquired on PLA-WSNT 0.05wt% sections subjected to varying wall shear stress (σ_w = 0.11, 0.15 and 0.21MPa at Ts = 130°C and t_s = 40s) along the (**A**) flow – vorticity plane (**v** – ∇x**v**), (**B**) vorticity – velocity gradient plane (∇x**v** – ∇**v**) and (**C**) flow – velocity gradient plane (**v** – ∇**v**). WAXS data are presented as (A-C, *left*) 2D patterns and (A-C, *right*) azimuthally [I(*q*)] and radially averaged [I(φ)] plots. Radial averaging for the I(φ) plots is performed in the vicinity of the (110)/(200) diffraction: $1.13 \leq q \leq 1.23$ Å^{-1}. The WAXS patterns are presented using a custom colormap with a scale that varies from 0 [white] to 1.75e^{-5} [red] counts.

Wide Angle X-ray scattering reveals orientation of crystallites in PLA-WSNT 0.1wt%

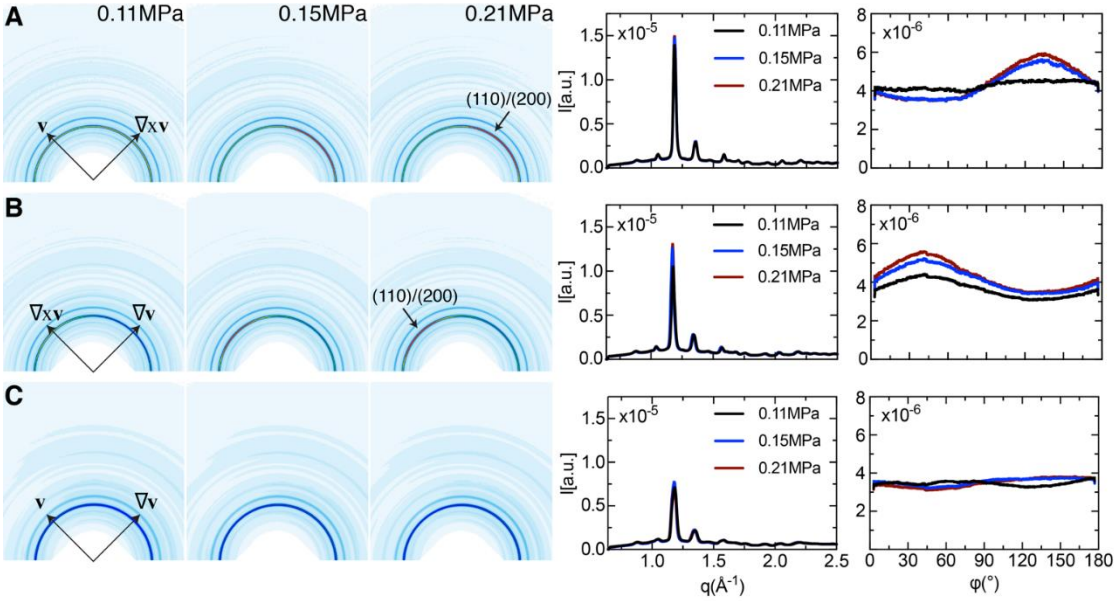

Figure S22. Wide angle X-ray scattering (WAXS) data were acquired on PLA-WSNT 0.1wt% sections subjected to varying wall shear stress (σ_w = 0.11, 0.15 and 0.21MPa at Ts = 130°C and t_s = 40s) along the (**A**) flow – vorticity plane (**v** – ∇x**v**), (**B**) vorticity – velocity gradient plane (∇x**v** – ∇**v**), and (**C**) flow – velocity gradient plane (**v** – ∇**v**). WAXS data are presented as (A-C, *left*) 2D patterns and (A-C, *right*) azimuthally [I(q)] and radially averaged [I(φ)] plots. Radial averaging for the I(φ) plots is performed in the vicinity of the (110)/(200) diffraction: $1.13 \leq q \leq 1.23$ Å^{-1}. The WAXS patterns are presented using a custom colormap with a scale that varies from 0 [white] to 1.75e^{-5} [red] counts.

References:

1. Son, Y. Determination of shear viscosity and shear rate from pressure drop an flow rate relationship in a rectangular channel. *Polymer* **48,** 632–637 (2007).
2. Aou, K., Kang, S. & Hsu, S. L. Morphological Study on Thermal Shrinkage and Dimensional Stability Associated with Oriented Poly (lactic acid). *Macromolecules* **38,** 7730–7735 (2005).
3. Pipe, C. J., Majmudar, T. S. & McKinley, G. H. High shear rate viscometry. *Rheol. Acta* **47,** 621–642 (2008).
4. Shribak, M. I. & Oldenbourg, R. Mapping polymer birefringence in 3D using a polarizing microscope with oblique illumination. *Proc. SPIE* **5462,** 57–67 (2004).
5. Yadgarov, L. *et al.* Strong light–matter interaction in tungsten disulfide nanotubes. *Phys. Chem. Chem. Phys.* 20812–20820 (2018). doi:10.1039/C8CP02245C
6. Garlotta, D. A Literature Review of Poly (Lactic Acid). *J. Polym. Environ.* **9,** (2002).
7. Fernandez-Ballester, L., Thurman, D. W. & Kornfield, J. A. Real-time depth sectioning: Isolating the effect of stress on structure development in pressure-driven flow. *J. Rheol. (N. Y. N. Y).* **53,** 1229 (2009).

Appendix A. Protocol and results from shotgun proteomics experiments

A.1 Introduction

BONCAT with pSILAC (introduced in **Chapter 1.3**) was used to probe the proteomic expression variation after peptide treatment in vascular endothelial cells. In one experiment, the cells were exposed to the integrin-binding RGD peptide (contains Arg-Gly-Asp integrin binding motif) and control RGE peptide (binds integrins at a much lower affinity[1]) to identify the proteomic changes after integrin modulation. In a second experiment, cells were treated with ALG-1001 therapeutic peptide and vehicle control in an effort to discover the proteomic regulation after drug exposure. In both studies, proteins synthesized during the peptide incubation period were enriched and analyzed using mass spectrometry to calculate the quantitative expression ratio between conditions. The following sections detail our experimental protocol as well as some observations from these two studies. This work was done in collaboration with Dr. Kai Yuet and Dr. Graham Hamblin, previous members of the Tirrell lab (Caltech).

A.2 RGD experiment

Reagents and methods

Pooled passage 2 (P2) Human Umbilical Vein Endothelial Cells (HUVECs) from American Type Culture Collection (ATCC) were cultured to P5 in a humidified incubator with pH controlled by 5% CO_2. The cells were grown in Vascular Cell Basal Medium (ATCC) with

Endothelial Cell Growth Kit-BBE (ATCC) and Antibiotic Antimycotic Solution (Sigma-Aldrich) added according to the manufacturer's instruction. The media was exchanged every two days. At P6, HUVECs were seeded at 5,000 cells per cm^2 density into eight T75 cell culture flasks (Greiner Bio-One) and cultured for three days in Endothelial Growth Medium (EGM, Lonza) with EGM Endothelial Cell Growth Medium SingleQuots (Lonza) and Antibiotic Antimycotic Solution added according to the manufacturer's instruction.

At end of the three day culture, the media were switched to BONCAT with pSILAC media (referred to as proteomics media) containing azidohomoalanine (AHA) and labeled amino acids. The proteomics media was based on a custom EGM media (Lonza) without manufacturer-added L-Arginine, L-Lysine and L-Methionine. The media was supplemented with EGM Endothelial Cell Growth Medium SingleQuots and Antibiotic Antimycotic Solution comparable to the standard EGM media except Fetal Bovine Serum (FBS) was replaced with dialyzed FBS (Thermo Fisher Scientific). L-Methionine (Sigma-Aldrich) was added to a final concentration of 33µM while AHA (Iris Biotech GmbH) was added to a final concentration of 1mM to achieve a 1:30 ratio (previously reported to be optimal for BONCAT experiments[2]). The media was also supplemented with either medium-labeled or heavy-labeled L-Lysine (medium: D_4 L-Lysine, heavy: $^{13}C_6{}^{15}N_2$ L-Lysine) and L-Arginine (medium: $^{13}C_6$ L-Arginine, heavy: $^{13}C_6{}^{15}N_4$ L-Arginine) sourced from Cambridge Isotope Laboratories. Labeled L-Lysine was added to a final concentration of 1mM while labeled L-Arginine was added to final concentration of 0.3mM. For four of the eight flasks, media were switched to the proteomics media with medium L-Lysine and L-Arginine while the other

four were switched to the proteomics media with heavy L-Lysine and L-Arginine. Concurrently, RGD peptide (sequence: GRGDTP, CPC Scientific) was added to two flasks with proteomics media containing medium-labeled amino acids and two flasks with proteomics media containing heavy-labeled amino acids. RGE (sequence: GRGESP, CPC Scientific) was added in similar manner to the remaining four flasks with both peptides added to a final concentration of 1mM.

Cells were incubated in the proteomics media for 24 hours and then detached using Typsin-EDTA solution (ATCC) and pelleted with a table top centrifuge (Beckman Coulter). We observed that at the end of the incubation, a fraction of the cells (estimate around 30%) exposed to the RGD peptide were detached from the flask and were not collected (detached cells likely undergo apoptosis with degraded proteome that may interfere with subsequent analysis). Cell lysis was performed with lysis buffer – PBS (Thermo Fisher) with 1% Sodium Dodecyl Sulfate (SDS, Sigma Aldrich), 100mM 2-chloroacetamide (Sigma Aldrich) and 1x protease inhibitor cocktail (Roche), at 90˚C for 10 minutes in the dark. Cellular debris was pelleted by centrifugation and the supernatant solution containing the total protein extract was collected for down-stream analysis. Total protein concentration was measured using Pierce BCA Protein Assay Kit (Thermo Fisher) followed by dilution of all samples to the same protein concentration using additional lysis buffer.

Next, replicates of RGD were individually paired with RGE samples and mixed in 1:1 ratio. In particular, medium-labeled samples were mixed with heavy-labeled samples to enable quantification of protein expression ratio in down-stream mass spectrometry analysis. After

mixing, 1/5th volume of PBS containing 0.8% SDS and 600mM 2-chloroacetamide (referred to as PBS/SDS/2Chl solution) was added to each sample and allowed to incubate for 30 minutes in the dark at 65˚C on a centrifuge tube shaker (Eppendorf). Meanwhile, 50uL per sample of DBCO agarose beads (Click Chemistry Tools) were washed three times with 10mL of PBS/SDS/2Chl solution. After 30 minutes of incubation on the shaker, the samples were then transferred to individual 15mL centrifuge tubes containing the washed DBCO agarose beads. Sample volume of PBS containing 8M urea, 0.85M NaCl, and 1x protease inhibitor was used to rinse the emptied sample tubes to collect any remaining sample solution and was added to the 15mL centrifuge tubes.

Samples were allowed to incubate with the beads for three hours covered in foil on a rotary mixer to facilitate click reaction enrichment of AHA-incorporated proteins. At end of the incubation step, 20µL of 100mM AHA was added for an additional 10 minute incubation on the rotary mixer. Afterwards, the samples were centrifuged and supernatant removed. Water was next added to wash the beads. Beads were then resuspended in 0.5mL of PBS/SDS/2Chl solution containing 1mM DTT (Sigma Aldrich) for 15 minute incubation at 70˚C, followed by 15 minutes cool down to room temperature. Supernatants were removed through centrifugation and beads incubated again with 0.5mL of PBS/SDS/2Chl solution containing 40mM iodoacetamide (Sigma Aldrich) for 30 minutes in the dark on the tube shaker.

The beads were next washed with 40mL of PBS/SDS/2Chl solution, 40mL of Tris (pH 8.0, Thermo Fisher) containing 8M urea (Sigma Aldrich) and 40mL of 20% acetonitrile in water (Sigma Aldrich). Washes were performed with 5mL of wash solution at a time and in the

order described above. After the washes, the beads were suspended in 200uL of 10% acetonitrile in 50mM ammonium bicarbonate (ammbic) and transferred to 1.7mL centrifuge tubes. Beads were then gently spun down and supernatant removed while keeping 100uL in each tube. On-bead digestion of the proteins was then performed by adding 10uL of 10% acetonitrile in 50mM ammbic containing 0.1µg of trypsin to each sample. The beads were allowed to incubate overnight on the shaker at 37˚C in the dark.

Following overnight trypsin digestion, the samples were added to filter columns with additional 20% acetonitrile in water used to ensure transfer of all samples. The solutions containing digested proteins were then collected by centrifugation and dried on a vacuum concentrator. Dried peptide pellets were then resuspended in 100uL of 50mM ammbic with a combination of vortex and sonication for 15 minutes. Next, HiPPR Detergent Removal Spin Column (Thermo Fisher) was used to remove detergent contaminants in the sample by following the manufacturer's instruction. Formic acid was then added to a final concentration of 0.2% to acidify the sample before sample concentration using C18 ZipTip (EMD Millipore) following the manufacturer's instruction. The sample was allowed to dry in the vacuum concentrator and then resuspended in 10uL of 0.2% formic acid with a combination of vortex and sonication. Finally, 5uL of each sample were analyzed using an EASY II nano-UPLC (Thermo Fisher) connected on-line to an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher) with two hours of run time per sample. All raw files obtained were then analyzed together in one *MaxQuant* search against human proteome obtained from

UniProt[3,4]. *MaxQuant* was set up to detect AHA and SILAC amino acids as well as with "re-quantification" and "match between runs" settings turned on.

Results and discussion

In this study, around 1200 to 1500 protein groups were identified in each sample. However, not all protein groups have a measurable heavy to medium-labeled protein ratio. Overall, 1092 protein groups were found to have at least one measured protein ratio with 439 groups contain ratio in all four samples. The protein groups with at least one protein ratio were submitted to *Limma* for differential expression analysis[5]. For each protein group, the leading protein with the most peptide evidence was chosen for analysis. In the rare cases where two or more protein groups have the same leading protein, the one with the most peptide evidence was retained. Results of *Limma* showed around 20% (229/1092) of the protein groups were identified to be differentially expressed (DE) with roughly 2/3rd of them up-regulated in the RGD-treated samples (**Figure A.1**).

Functional analysis of the DE proteins was performed using *DAVID* with the up- and down-regulated proteins analyzed separately[6]. All discovered gene ontology (GO) biological processes were visualized using *REVIGO*, with similarity level set to medium and GO annotation size determined using the *Uniport Homo Sapiens* database[4,7]. GO biological processes with FDR<0.05 were considered to be statistically significant. With both the up- and down-regulated protein lists, "cell-cell adhesion" was the only biological process found to be strongly enriched (**Figure A.2**).

In context of the natural roles of RGD-containing proteins and their targeted integrins in cell adhesion and recognition[8], modulation of proteins associated with "cell-cell adhesion" after integrin binding is not unexpected. During RGD exposure, binding of the peptide with cell surface integrins likely activated a diverse group of downstream pathways that regulated adhesion-related protein expression[9]. Previous microarray studies that probed RGD-induced transcriptome regulation found genes in immune response, hematological system development, cell migration, development, growth, and death are modulated[10–12]. In our proteomics study, we were unable to validate many of these previous findings, possibly due to the differences in throughput of the technologies: microarray is able to measure expression changes across tens of thousands of genes, while our proteomics approach quantified only around one thousand proteins. Nevertheless, modulation of "cell-cell adhesion" proteins is consistent with known mechanism of the RGD peptide.

A.3 ALG-1001 experiment

Reagents and methods

The BONCAT with pSILAC experiment with RGD and RGE demonstrated that proteomics can be used to identify proteome perturbation during peptide treatment. Next, we applied the technique to identify proteomic changes during ALG-1001 exposure. Between the two studies, some additional improvements were made to the proteomics sample preparation protocol with the updated protocol shown below.

Pooled P2 Human Umbilical Vein Endothelial Cells (HUVECs) from American Type Culture Collection (ATCC) were cultured to P4 in a humidified incubator with pH controlled by 5% CO_2. The cells were grown in Vascular Cell Basal Medium (ATCC) with Endothelial Cell Growth Kit-BBE (ATCC) and Antibiotic Antimycotic Solution (Sigma-Aldrich) added according to the manufacturer's instruction. The media was exchanged every two days. At P5, HUVECs were seeded at 4,000 cells per cm^2 density into twelve T75 cell culture flasks (Greiner Bio-One) and cultured for two days in Endothelial Growth Medium (EGM, Lonza) with EGM Endothelial Cell Growth Medium SingleQuots (Lonza) and Antibiotic Antimycotic Solution added according to the manufacturer's instruction.

At end of two day culture, the media were switched to BONCAT with pSILAC media (referred to as proteomics media) containing AHA and labeled amino acids. The proteomics media was based on a custom EGM media (Lonza) without manufacturer-added L-Arginine, L-Lysine and L-Methionine. The media was supplemented with EGM Endothelial Cell Growth Medium SingleQuots and Antibiotic Antimycotic Solution comparable to the standard EGM media except FBS was replaced with dialyzed FBS (Thermo Fisher Scientific). L-Methionine (Sigma-Aldrich) was added to a final concentration of 33µM while AHA (Iris Biotech GmbH) was added to a final concentration of 1mM to achieve a 1:30 ratio (previously reported to be optimal for BONCAT experiments[2]). The media was also supplemented with either medium-labeled or only heavy-labeled L-Lysine (medium: D_4 L-Lysine, heavy: $^{13}C_6{}^{15}N_2$ L-Lysine) and L-Arginine (medium: $^{13}C_6$ L-Arginine, heavy: $^{13}C_6{}^{15}N_4$ L-Arginine) sourced from Cambridge Isotope Laboratories. Labeled L-Lysine was

added to a final concentration of 1mM while labeled L-Arginine was added to a final concentration of 0.3mM. For six of the twelve flasks, media were switched to the proteomics media with medium L-Lysine and L-Arginine while the other six were switched to the proteomics media with heavy L-Lysine and L-Arginine. Concurrently, concentrated ALG-1001 peptide (CPC Scientific) dissolved in DI water was added to a final concentration of 100uM in three flasks with proteomics media containing medium-labeled amino acids and three flasks with proteomics media containing heavy-labeled amino acids. An equal volume of sterilized DI water was added in similar manner to the remaining six flasks to serve as vehicle control.

Cells were incubated in the proteomics media for 48 hours and then detached using Typsin-EDTA solution (ATCC) and pelleted with a table top centrifuge (Beckman Coulter). Cell lysis was performed with lysis buffer – PBS (Thermo Fisher) with 1% Sodium Dodecyl Sulfate (Sigma Aldrich), 100mM 2-chloroacetamide (Sigma Aldrich) and 1x protease inhibitor cocktail (Roche), at 90˚C for 10 minutes in the dark. Cellular debris was pelleted by centrifugation and the supernatant solution containing the total protein extract was collected for down-stream analysis. Total protein concentration was measured using Pierce BCA Protein Assay Kit (Thermo Fisher) followed by dilution of all samples to the same protein concentration using additional lysis buffer.

Next, replicates of ALG-1001 treated samples were paired with control samples and mixed in 1:1 ratio to form six samples. In particular, medium-labeled samples were mixed with heavy-labeled samples to enable quantification of protein expression ratio in down-stream

mass spectrometry analysis. After mixing, 1/5th volume of PBS containing 0.8% SDS and 600mM 2-chloroacetamide (subsequently referred to as PBS/SDS/2Chl solution) was added to each sample and allowed to incubate for 30 minutes in the dark at 65˚C on a centrifuge tube shaker (Eppendorf). Meanwhile, 50uL per sample of DBCO agarose beads (Click Chemistry Tools) were washed with 10mL of PBS/SDS/2Chl solution, 10mL of Tris (pH 8.0, Thermo Fisher) containing 8M urea (Sigma Aldrich), 10mL of 20% acetonitrile in water and 20mL of PBS/SDS/2Chl solution. After 30 minutes incubation on the shaker, the samples were then transferred to individual 15mL centrifuge tubes containing the washed DBCO agarose beads. Sample volume of PBS containing 8M urea, 0.85M NaCl and 1x protease inhibitor was used to rinse the emptied sample tubes to collect any remaining sample solution and are then added to the 15mL centrifuge tubes. Samples were incubated for three hours covered in foil on a rotary mixer to facilitate click reaction enrichment of AHA-incorporated proteins. Next, sample volume of PBS containing 8M urea, 0.85M NaCl and 1x protease inhibitor was added and the samples were allowed to incubate overnight in foil on the rotary mixer. The additional incubation step helps improve the amount of AHA-containing proteins recovered.

At end of incubation, 30µL of 100mM AHA was added for additional 10 minute incubation on the rotary mixer. Afterwards, the samples were centrifuged and the supernatant removed. Water was next added to wash the beads and removed through centrifugation. Beads were then resuspended in 0.5mL of PBS/SDS/2Chl solution containing 1mM DTT (Sigma Aldrich) for 15 minute incubation at 70˚C, followed by 15 minutes cool down to room

temperature. Supernatants were removed through centrifugation and beads incubated again with 0.5mL of PBS/SDS/2Chl solution containing 40mM iodoacetamide (Sigma Aldrich) for 30 minutes in the dark on the tube shaker.

The samples were next washed with 40mL of PBS/SDS/2Chl solution, 40mL of Tris (pH 8.0, Thermo Fisher) containing 8M urea (Sigma Aldrich) and 40mL of 20% acetonitrile in water (Sigma Aldrich). Washes were performed with 5mL of wash solution at a time and in the order described above. After the washes, the beads were suspended in 200uL of 10% acetonitrile in 50mM ammonium bicarbonate (ammbic) and transferred to 1.7mL centrifuge tubes. Beads were then gently spin down and supernatant removed while keeping 150uL in each tube. On bead digestion of proteins was then performed by adding 10uL of 10% acetonitrile in 50mM ammbic containing 0.2µg of Lys-C protease to each sample. The beads were allowed to incubate for two hours on shaker at 37˚C in the dark. Next, 10uL of 10% acetonitrile in 50mM ammbic containing 0.2µg of trypsin was added to each sample and the samples allowed to incubate overnight on the shaker at 37˚C in the dark.

After overnight trypsin digestion, the samples were added to filter columns and 20% acetonitrile in water was used to ensure transfer of all samples. The solution containing digested proteins was collected by centrifugation and dried on a vacuum concentrator. Dried peptide pellets were then resuspended in 100uL of 50mM ammbic with a combination of vortex and sonication for 15 minutes. Next, HiPPR Detergent Removal Spin Column (Thermo Fisher) was used to remove detergent contaminants in the sample by following the manufacturer’s instruction.

After detergent cleanup, vacuum centrifuge was used to reduce sample volume to around 20uL. Formic acid was then added to a final concentration of 0.2% to acidify the sample before sample concentration using C18 ZipTip (EMD Millipore) following the manufacturer's instruction. ZipTip was performed twice to improve recovery of samples. The samples were then dried in vacuum concentrator and then resuspended in 20uL of 0.2% formic acid with a combination of vortex and sonication. Finally, 1uL of each sample was analyzed using an EASY II nano-UPLC (Thermo Fisher) connected on-line to an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher) with two hours of run time per sample. All raw files obtained were then analyzed together in one *MaxQuant* search against human proteome obtained from UniProt[3,4]. *MaxQuant* was set up to detect AHA and SILAC amino acids as well as with "re-quantification" and "match between runs" settings turned on.

Among the six samples tested, detection doubled to around 2700 to 2800 protein groups per sample, due to several improvements in sample preparation. One major improvement potentially came from the extended (48 hours in this study vs. 24 hours in RGD vs. RGE study) proteomics media exposure that ensured there is a larger fraction of AHA-incorporated proteome for enrichment. Next, bead-based click reaction was optimized to improve overall efficiency and was allowed to react overnight to enhance enrichment of newly synthesized proteins. After that, both Lys-C and trypsin proteases were used to enhance the digestion of proteins into expected peptide fragments. Lastly, the ZipTip desalting step was performed twice to increase the amount of peptide recovery. Overall,

implementation of these technical changes significantly improved protein group detection in mass spectrometry.

Results and discussion

Among the protein groups discovered, 2440 were found to have at least one measured protein ratio and 1141 groups that contained ratios in all six samples. The protein groups with at least one protein ratio were submitted to *Limma* for differential expression analysis[5]. For each protein group, the leading protein with the most peptide evidence associated was chosen for analysis. In the rare cases where two or more protein groups have the same leading protein, the one with the most peptide evidence was retained.

Limma results showed that no detectable proteins was found to be differentially expressed with statistical significance (**Figure A.3**). After further analysis, we found that compared to the RGD vs. RGE study, a larger fraction of proteins have a close to zero fold change in this study (**Figure A.4**). Additionally, this pattern is consistent with two previous RNA-seq studies that found perturbation by RGD is significantly stronger than ALG-1001 (Kornfield group, Caltech, unpublished). On the other hand, the two proteomics studies presented here are not directly comparable, since many experimental parameters and sample preparation steps were changed. Experimentally, treatment time was doubled (48 hours in this study vs. 24 hours in the RGD vs. RGE study), while the peptide concentration was reduced by 10 folds. The lower concentration was used to reduce effects of non-specific binding, but this dosage may be too mild to elicit detectable proteomic changes. Thus, these findings do not demonstrate the proteome is not regulated after ALG-1001 treatment, but that the effect on

the proteome may be too small to be detected at the dosage tested. To improve discovery, higher dosages of ALG-1001 and more biological replicates should be employed.

A.4 Figures and tables

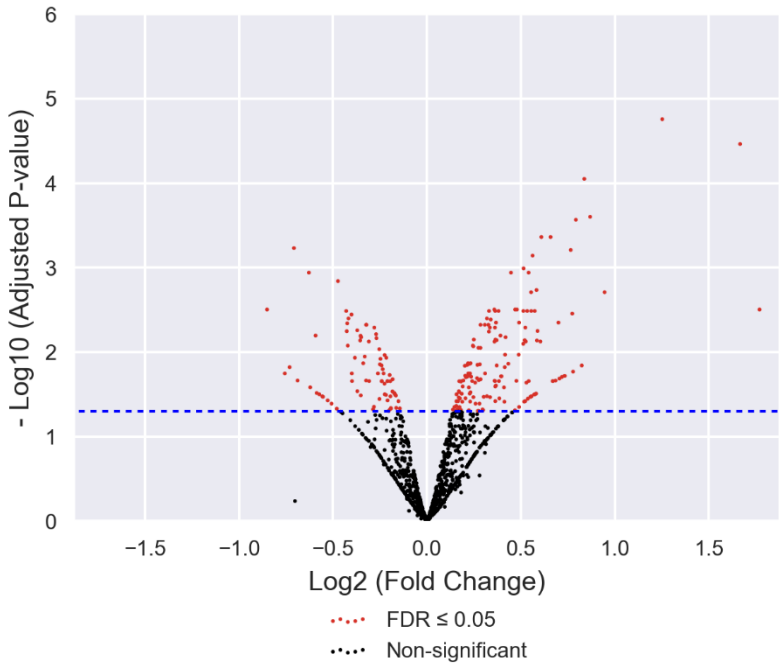

Figure A.1 Volcano plot of *Limma* result containing proteins regulated between RGD and RGE treatments

Limma-measured fold change and adjusted p-values were plotted with each dot representing a protein. Proteins colored red are considered DE by an adjusted p-value (FDR) cutoff of 0.05 (blue dotted line).

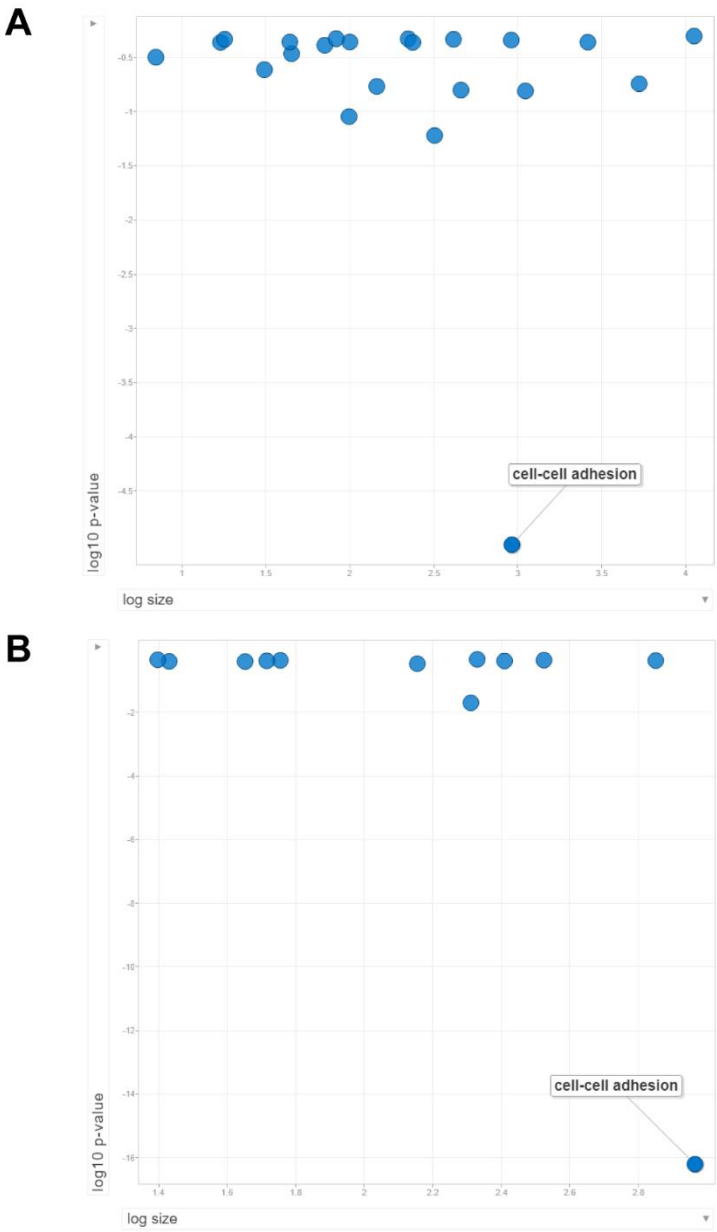

Figure A.2 *REVIGO* plot of GO biological processes-enriched with DE genes between RGD and RGE treatments

Up- (**A**) and down-regulated (**B**) DE proteins were submitted for GO enrichment with results visualized using *REVIGO*. P-value is *REVIGO*-derived GO enrichment P-value. Size is the number of human proteins associated with each GO term. Strongly enriched processes are labeled.

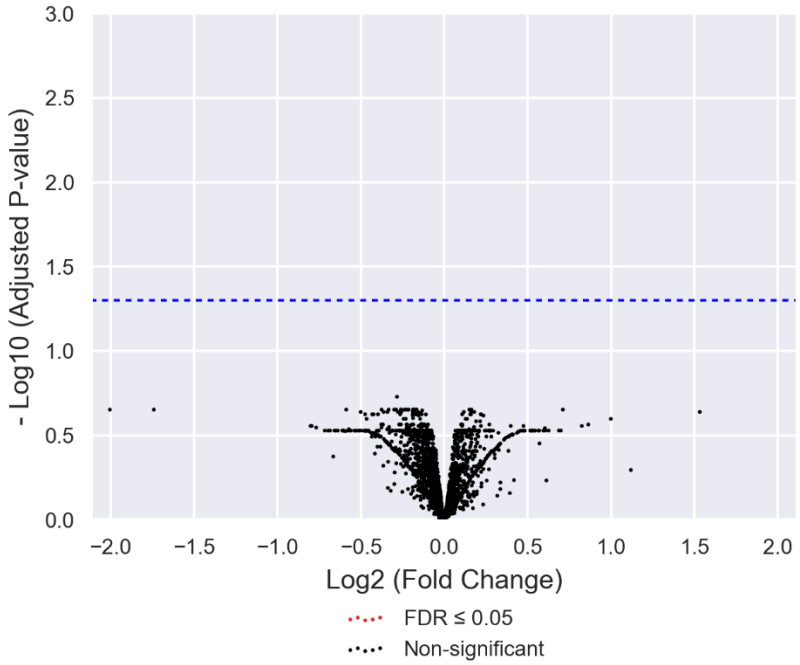

Figure A.3 Volcano plot of *Limma* results shows no protein regulated by ALG-1001 with statistical significance

Limma-measured fold change and adjusted p-values are plotted with each dot representing a protein. No DE proteins are detected with an adjusted p-value (FDR) cutoff of 0.05 (blue dotted line).

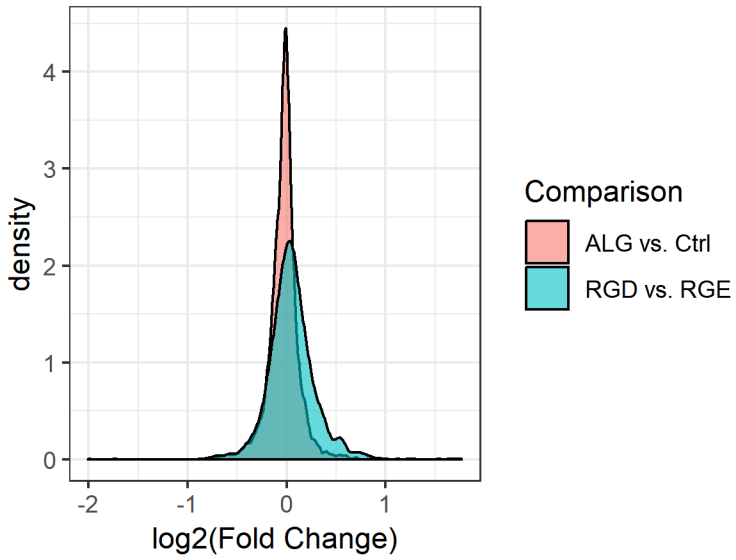

Figure A.4 Density plot of the fold changes measured among the two comparisons
Fold changes calculated by *Limma* are plotted in form of density plot. Area under the curve sums to 1 for each of the comparisons. ALG vs. Ctrl is ALG-1001 vs. vehicle control comparison, which shows more proteins with near zero fold change.

A.5 References

1. Patterson, K. C. *et al.* Measurement of Cationic and Intracellular Modulation of Integrin Binding Affinity by AFM-Based Nanorobot. *Biophys. J.* **105**, 40–47 (2013).
2. Bagert, J. D. *et al.* Quantitative, Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labeling by Amino Acids in Cell Culture. *Mol. Cell. Proteomics* **13**, 1352–1358 (2014).
3. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
4. Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
5. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* gkv007 (2015). doi:10.1093/nar/gkv007
6. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2008).
7. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).
8. Barczyk, M., Carracedo, S. & Gullberg, D. Integrins. *Cell Tissue Res.* **339**, 269–280 (2010).
9. Ruegg, C. & Mariotti, A. Vascular integrins: pleiotropic adhesion and signaling molecules in vascular homeostasis and angiogenesis. *Cell. Mol. Life Sci.* **60**, 1135–1157 (2003).
10. Bretschi, M. *et al.* Cilengitide affects tumor compartment, vascularization and microenvironment in experimental bone metastases as shown by longitudinal 18F-FDG PET and gene expression analysis. *J. Cancer Res. Clin. Oncol.* **139**, 573–583 (2013).
11. Dia, V. P. & Gonzalez de Mejia, E. Differential gene expression of RAW 264.7 macrophages in response to the RGD peptide lunasin with and without lipopolysaccharide stimulation. *Peptides* **32**, 1979–1988 (2011).
12. Huang, T.-C. *et al.* An apoptosis-related gene network induced by novel compound-cRGD in human breast cancer cells. *FEBS Lett.* **581**, 3517–3522 (2007).